YOKOHAMA MATHEMATICAL
JOURNAL VOL. 40, 1992

INEQUALITIES FOR THE r-METRIC

By

A. G. M. STEERNEMAN

(Received March 6, 1991)

Summary. Matusita's r-metric on the set of probability measures on a measurable space is extended to the set of signed measures. Special cases are $r=1$ and $r=2$ which coincide with the total variation and Hellinger metric respectively. Inequalities between these metrics are presented. As a consequence it is obtained that the topologies induced by the total variation norm and the r-metric are equivalent.

1. Introduction

In statistics and probability theory frequent use is made of two well-known metrics on the space $\mathcal{M}_1(\mathcal{X}, \mathcal{F})$ of probability measures on the measurable space $(\mathcal{X}, \mathcal{F})$, namely the total variation metric d_v and the Hellinger metric d_H. These metrics are used for instance in Kakutani [6], Kraft [7], LeCam [8], Dacunha-Castelle [3], and Reiss [12]. The total variation and Hellinger metric are obtained as special cases of the so-called r-metric d_r taking $r=1$ and $r=2$, respectively. The metric d_r plays a nice part in the study of the affinity of several probability distributions. These notions were introduced by Matusita [9, 10].

If the probability measures P and Q on $(\mathcal{X}, \mathcal{F})$ are dominated by the σ-finite measure λ, and p and q respectively are nonnegative versions of the Radon-Nikodym derivatives of P and Q with respect to λ, then

$$d_r(P, Q)=\left[\int_{\mathcal{X}} |p^{1/r}-q^{1/r}|^r d\lambda\right]^{1/r}. \tag{1.1}$$

In fact Matusita [9] uses this metric only for $r=1, 2, \cdots$. We have

$$d_v(P, Q)=d_1(P, Q)=\int_{\mathcal{X}} |p-q|\, d\lambda \tag{1.2}$$

and

$$d_H(P, Q)=d_2(P, Q)=\sqrt{2-2\rho(P, Q)}, \tag{1.3}$$

where

$$\rho(P, Q)=\int_{\mathcal{X}} p^{1/2}q^{1/2}d\lambda, \tag{1.4}$$

1991 Mathematics Subject Classification: Primary 28A33, Secondary 60A10.
Key words and phrases: r-metric, total variation norm, Hellinger affinity and metric.

which is called the Hellinger affinity of P and Q. Kraft [7] established the following fundamental inequality

$$d_H^2(P, Q) \leqq d_v(P, Q) \leqq 2(1-\rho^2(P, Q))^{1/2} \leqq 2d_H(P, Q). \tag{1.5}$$

The Hellinger affinity defined in (1.4) is a nice tool in deriving e.g. an inequality needed for estimating the total variation distance between two product measures in terms of distances between the separate components. For $i=1, \cdots, k$ let $(\mathcal{X}_i, \mathcal{F}_i)$ be a measurable space and P_i and Q_i probability measures on $(\mathcal{X}_i, \mathcal{F}_i)$, then

$$\begin{aligned}2-2\exp(-8^{-1}\textstyle\sum_{i=1}^k d_v^2(P_i, Q_i)) &\leqq d_v(\times_{i=1}^k P_i, \times_{i=1}^k Q_i)\\ &\leqq \textstyle\sum_{i=1}^k d_v(P_i, Q_i).\end{aligned} \tag{1.6}$$

With regard to this inequality reference should be made to [5], [2], [1], [13], and [12]. Inequalities like (1.6) can be applied in order to derive so-called zero-one laws or equivalence-orthogonality dichotomies of products of probability measures. We refer to e.g. [6], [2], [11], [13], and [4].

In order to obtain an asymptotic expansion of the joint distribution of order statistics Reiss [12] needed an inequality like (1.6) for finite signed measures:

$$d_v(\times_{i=1}^k \mu_i, \times_{i=1}^k \nu_i) \leqq \textstyle\sum_{i=1}^k (\prod_{j=1}^{i-1}\|\mu_j\|)(\prod_{j=i+1}^k \|\nu_j\|) d_v(\mu_i, \nu_i), \tag{1.7}$$

where μ_i and ν_i are finite signed measures on the measurable space $(\mathcal{X}_i, \mathcal{F}_i)$, $\|\mu_i\| = d_v(\mu_i, 0)$ and $\|\nu_j\| = d_v(\nu_j, 0)$. (Products in (1.7) with empty index-set are defined to be equal to one.) The question arose whether a lower bound could be obtained for the total variation distance between two products of finite signed measures, having (1.6) as a special case. Note that the right part of (1.6) follows from (1.7). Since the derivation of (1.6) heavily depends on (1.3) and (1.5), a first task is to extend the concepts Hellinger affinity and metric to finite signed measures. This is done in [14] with the following consequence:

$$\begin{aligned}d_v(\times_{i=1}^k \mu_i, \times_{i=1}^k \nu_i) \geqq{}& \textstyle\prod_{i=1}^k \|\mu_i\| + \prod_{i=1}^k \|\nu_i\| +\\ &-2[\textstyle\prod_{i=1}^k \|\mu_i\| \cdot \|\nu_i\|]^{1/2} \exp(-8^{-1}\sum_{i=1}^k d_v^2(\tilde{\mu}_i, \tilde{\nu}_i)),\end{aligned} \tag{1.8}$$

where $\tilde{\mu}_i = \|\mu_i\|^{-1}|\mu_i|$ and $\tilde{\nu}_i = \|\nu_i\|^{-1}|\nu_i|$ for $i=1, \cdots, k$. Recall that $|\mu_i|$ and $|\nu_i|$ are the total variation measures associated with μ_i and ν_i.

Another question arises. Can Matusita's r-metric be extended to finite signed measures for arbitrary $r \geqq 1$? (Matusita [9] considered $r=1, 2, \cdots$ for probability measures.) In section 2 we answer this question affirmatively. Section 3 presents extensions of some results from [9]. The proofs are simpler for the cases where Matusita assumes that r is a natural number.

2. The r-metric for signed measures

Let $\mathcal{M}=\mathcal{M}(\mathcal{X}, \mathcal{F})$ denote the space of finite signed measures on the measurable space $(\mathcal{X}, \mathcal{F})$, and $\mathcal{M}_1$ the subset of all probability measures. For $\tau\in\mathcal{M}$ let $|\tau|$ denote the total variation measure associated with τ. $\mathcal{M}$ is a Banach lattice with respect to the total variation norm $\|\cdot\|$ defined by $\|\tau\|=|\tau|(\mathcal{X})$. This norm has the extra property that $\|\mu+\nu\|=\|\mu\|+\|\nu\|$ for $\mu, \nu\geqq 0$ (see [15], 369-370). The total variation metric d_v on $\mathcal{M}$ is the metric induced by this norm. For $\mu, \nu\in\mathcal{M}$, let λ be a σ-finite measure on $(\mathcal{X}, \mathcal{F})$ such that $\mu, \nu\ll\lambda$ and let $f\in d\mu/d\lambda$, $g\in d\nu/d\lambda$, then

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

Matusita [9] introduced the r-metric for probability measures. We extend Matusita's definition to arbitrary finite signed measures and generalize the restriction $r=1, 2, \cdots$ to $r\geqq 1$. This is done by defining

$$d_r(\mu, \nu)=\left[\int_{\mathcal{X}}|f_+^{1/r}-g_+^{1/r}|^r d\lambda+\int_{\mathcal{X}}|f_-^{1/r}-g_-^{1/r}|^r d\lambda\right]^{1/r}, \tag{2.2}$$

where $x_+=\max(x, 0)$ and $x_-=-\min(x, 0)$. Note that the definition of $d_r(\mu, \nu)$ does not depend on the choices of λ and $f\in d\mu/d\lambda$, $g\in d\nu/d\lambda$. Since $f, g\in\mathcal{L}_1$, we have that $f_+^{1/r}, g_+^{1/r}, f_-^{1/r}, g_-^{1/r}\in\mathcal{L}_r$. Hence $|f_+^{1/r}-g_+^{1/r}|, |f_-^{1/r}-g_-^{1/r}|\in\mathcal{L}_r$ and thus $d_r(\mu, \nu)$ is well-defined.

Let $\|\cdot\|_r$ denote the r-norm on $\boldsymbol{R}^2$: $\|(x, y)\|_r=(|x|^r+|y|^r)^{1/r}$. Then

$$d_r(\mu, \nu)=\|(d_r(\mu_+, \nu_+), d_r(\mu_-, \nu_-))\|_r, \tag{2.3}$$

where $\mu_+(B)=\int_B f_+ d\lambda$ for any $B\in\mathcal{F}$, etc. In establishing that d_r is a metric, the triangle inequality follows from (2.3) and Minkowski's inequality:

$$\begin{aligned}d_r(\mu, \tau)&\leqq\|(d_r(\mu_+, \nu_+)+d_r(\nu_+, \tau_+), d_r(\mu_-, \nu_-)+d_r(\nu_-, \tau_-))\|_r\\&\leqq\|(d_r(\mu_+, \nu_+), d_r(\mu_-, \nu_-))\|_r+\|(d_r(\nu_+, \tau_+), d_r(\nu_-, \tau_-))\|_r\\&=d_r(\mu, \nu)+d_r(\nu, \tau).\end{aligned}$$

For $r=1$ we obtain the total variation metric d_v. Taking $r=2$ the Hellinger metric d_H is found. For probability measures this metric is well-known, see e.g. [6]. In [14] it was generalized to finite signed measures. It can easily be seen that

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

$$\rho(\mu, \nu)=\int_{\mathcal{X}}f_+^{1/2}g_+^{1/2}d\lambda+\int_{\mathcal{X}}f_-^{1/2}g_-^{1/2}d\lambda=\int_A|fg|^{1/2}d\lambda, \tag{2.5}$$

where $A=\{fg>0\}$. The quantity $\rho(\mu, \nu)$ is called the Hellinger affinity between

μ and ν. Matusita [9] and [10] consider affinities between more than two probability measures.

3. Inequalities for the r-metric

The total variation distance between two finite signed measures can be estimated by using r-metrics. We need the following lemma which is an immediate consequence of Hölder's inequality.

Lemma 3.1. *Let $a, b, c, d \geqq 0$. If $r, s > 1$ with $r^{-1}+s^{-1}=1$, then $a^{1/r}b^{1/s}+c^{1/r}d^{1/s} \leqq (a+c)^{1/r}(b+d)^{1/s}$.*

The next theorem extends the results contained in the Theorems 4 and 5 of [9]. The upper bound presented is believed to be new.

Theorem 3.2. *Let $\mu, \nu \in \mathcal{M}$, then $d_r^r(\mu, \nu)$ is decreasing in r on $[1, \infty)$. If $r, s>1$ with $r^{-1}+s^{-1}=1$, then*

$$d_r^r(\mu, \nu) \leqq d_v(\mu, \nu) \leqq \|\mu\|^{1/s} d_r(\mu, \nu) + \|\nu\|^{1/r} d_s(\mu, \nu). \tag{3.1}$$

Proof. Note that the function $\varphi(x)=|a^{1/x}-b^{1/x}|^x$, where $a, b>0$, is decreasing on $(0, \infty)$. Using this fact, formula (2.2) and the relation $|x-y|=|x_+-y_+| + |x_- - y_-|$, $x, y \in \boldsymbol{R}$, we obtain the first inequality of (3.1). The second one is established as follows. By applying Hölder's inequality we derive

$$\begin{aligned} d_v(\mu_+, \nu_+) &= \int |f_+ - g_+| \, d\lambda \\ &\leqq \int |f_+^{1/r} - g_+^{1/r}| f_+^{1/s} d\lambda + \int |f_+^{1/s} - g_+^{1/s}| g_+^{1/r} d\lambda \\ &\leqq \|\mu_+\|^{1/s} d_r(\mu_+, \nu_+) + \|\nu_+\|^{1/r} d_s(\mu_+, \nu_+). \end{aligned}$$

Analogously we obtain

$$d_v(\mu_-, \nu_-) \leqq \|\mu_-\|^{1/s} d_r(\mu_-, \nu_-) + \|\mu_-\|^{1/r} d_s(\mu_-, \nu_-).$$

Lemma 3.1 provides

$$\begin{aligned} d_v(\mu, \nu) &= d_v(\mu_+, \nu_+) + d_v(\mu_-, \nu_-) \\ &\leqq (\|\mu_+\| + \|\mu_-\|)^{1/s} (d_r^r(\mu_+, \nu_+) + d_r^r(\mu_-, \nu_-))^{1/r} \\ &\quad + (\|\nu_+\| + \|\nu_-\|)^{1/r} (d_s^s(\mu_+, \nu_+) + d_s^s(\mu_-, \nu_-))^{1/s} \\ &= \|\mu\|^{1/s} d_r(\mu, \nu) + \|\nu\|^{1/r} d_s(\mu, \nu). \quad \square \end{aligned}$$

Taking $r=2$ we obtain from Theorem 2 an upper and lower bound of d_v in terms of d_H. However, in this special case the upper bound can be sharpened, by using the Hellinger affinity defined in (2.5):

$$d_v(\mu, \nu) \leqq [(\|\mu\| + \|\nu\|)^2 - 4\rho^2(\mu, \nu)]^{1/2} \leqq (\|\mu\|^{1/2} + \|\nu\|^{1/2}) d_H(\mu, \nu).$$

This result is an extension of Lemma 1 of [7] and it is established in [14].

The following result provides another upper estimate for the total variation distance in terms of only one r-metric instead of two as was the case in Theorem 3.2. It generalizes Matusita [9], Theorem 3, which contains the result for probability measures in the case that $r>1$ is a natural number.

Theorem 3.3. *For $r, s>1$ with $r^{-1}+s^{-1}=1$,*

$$d_v(\mu, \nu) \leqq \max(1, r/2)(\|\mu\|^{1/s} + \|\nu\|^{1/s}) d_r(\mu, \nu). \tag{3.7}$$

Proof. Define $C=\max(1, r/2)$. Using the inequality

$$t^r - 1 \leqq C(t-1)(t^{r-1}+1), \qquad t \geqq 1,$$

we obtain according to Lemma 3.1 that

$$\begin{aligned} |f-g| &= |f_+ - g_+| + |f_- - g_-| \\ &\leqq C\{|f_+^{1/r} - g_+^{1/r}|(f_+^{1/s} + g_+^{1/s}) + |f_-^{1/r} - g_-^{1/r}|(f_-^{1/s} + g_-^{1/s})\} \\ &\leqq C(|f_+^{1/r} - g_+^{1/r}|^r + |f_-^{1/r} - g_-^{1/r}|^r)^{1/r} \\ &\quad \times ((f_+^{1/s} + g_+^{1/s})^s + (f_-^{1/s} + g_-^{1/s})^s)^{1/s} \\ &\leqq C(|f_+^{1/r} - g_+^{1/r}|^r + |f_-^{1/r} - g_-^{1/r}|^r)^{1/r}(|f|^{1/s} + |g|^{1/s}). \end{aligned}$$

The result follows by applying the inequalities of Hölder and Minkowski.

The following result is suggested by Prof. Dr. A. J. Stam. It follows from the Theorems 3.2 and 3.3.

Theorem 3.4. *The topologies on $\mathcal{M}$ induced by $\|\cdot\|$ resp. $d_r (r>1)$ are equivalent.*

Proof. Let s satisfy the relation $r^{-1}+s^{-1}=1$. Let $\mu \in \mathcal{M}$ and consider the sequence $\{\mu_n\}$ in $\mathcal{M}$. If $d_v(\mu_n, \mu) \to 0$, then according to Theorem 3.2 it follows that $d_r(\mu_n, \mu) \to 0$. Conversely suppose that $d_r(\mu_n, \mu) \to 0$. Note that $\|\mu_n\| = d_r^r(\mu_n, 0)$, hence the sequence $\{\|\mu_n\|\}$ is bounded. From Theorem 3.3 it now follows that $d_v(\mu_n, \mu) \to 0$. □

Acknowledgement. I would like to thank Dr. A. E. Ronner, Prof. Dr. W. Schaafsma and Prof. Dr. A. J. Stam for their careful reading of an earlier version of the manuscript.

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] D. Dacunha-Castelle, Vitesse de convergence pour certains problèmes statistique, *Lecture Notes in Mathematics*, **678** (1978), Springer-Verlag, Berlin-Heidelberg-New York.

[4] A. Hillion, Sur l'intégrale Hellinger et la séparation asymptotique, *C.R. Acad. Sc. Paris*, **283** (1976), ser. A, 61-64.

[5] W. Hoeffding and J. Wolfowitz, Distinguishability of sets of distributions, *Ann. Math. Statist.*, **29** (1958), 700-718.

[6] S. Kakutani, On equivalence of infinite product measure, *Ann. Math.*, **49** (1948), 214-224.

[7] C.H. Kraft, Some conditions for consistency and uniform consistency of statistical procedures, *Univ. California Publications in Statist.*, **2** (1955), 125-141.

[8] L. LeCam, Convergence of estimates under dimensionality restrictions, *Ann. Statist.*, **1** (1973), 38-53.

[9] K. Matusita, On the notion of affinity of several distributions and some of its applications, *Ann. Inst. Statist. Math.*, **19** (1967), 181-192.

[10] K. Matusita, Some properties of affinity and applications, *Ann. Inst. Statist. Math.*, **23**, (1971), 137-155.

[11] T. Nemetz, Equivalence-ortogonality dichotomies of probability measures, *Coll. Math. Soc. János Bolyai*, **11** (1974).

[12] R.D. Reiss, Approximation of product measures with an application to order statistics, *Ann. Probab.*, **9** (1981), 335-341.

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

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

[15] K. Yosida, *Functional analysis*, 5th ed., Springer-Verlag, Berlin-Heidelberg-New York.

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