Adv. Appl. Prob. **29**, 429–443 (1997)
Printed in N. Ireland

INTEGRAL PROBABILITY METRICS AND THEIR GENERATING CLASSES OF FUNCTIONS

ALFRED MÜLLER,* *Universität Karlsruhe*

Abstract

We consider probability metrics of the following type: for a class $\mathfrak{F}$ of functions and probability measures P, Q we define $d_{\mathfrak{F}}(P, Q) := \sup_{f\in\mathfrak{F}} |\int f\,dP - \int f\,dQ|$. A unified study of such *integral probability metrics* is given. We characterize the maximal class of functions that generates such a metric. Further, we show how some interesting properties of these probability metrics arise directly from conditions on the generating class of functions. The results are illustrated by several examples, including the Kolmogorov metric, the Dudley metric and the stop-loss metric.

INTEGRAL PROBABILITY METRICS; MAXIMAL GENERATOR; UNIFORMITY IN WEAK CONVERGENCE: STOP-LOSS METRIC.

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60E05
SECONDARY 60B10

1. Introduction

Many models in applied probability are so complex that an explicit calculation of their characteristics is impossible. Therefore approximations are of practical importance. But these approximations require some sort of stability of the model. Very often it is convenient to express stability in terms of probability metrics. Since most of the characteristics are defined as an integral of some function f with respect to a probability measure P, probability metrics based on the comparison of integrals are of special interest.

In recent years there has appeared a vast literature on the theory of probability metrics. Many results have been summarized in the monograph by Rachev (1991). Many of the numerous metrics, which have proved to be useful, are based on the comparison of integrals as follows. For a class $\mathfrak{F}$ of functions let

$$d(P, Q) := \sup_{f\in\mathfrak{F}} \left| \int f\,dP - \int f\,dQ \right|.$$

In Zolotarev (1983) these metrics are called *probability metrics with a ζ-structure.* We will use the more intuitive terminology *integral probability metric.*

The purpose of this paper is to give a unified study of integral probability metrics. After providing some preliminary results from functional analysis in Section 2, we

Received 17 July 1995; revision received 20 November 1995.
* Postal address: Institut für Wirtschaftstheorie und Operations Research, Universität Karlsruhe, Kaiserstr. 12, D-76128 Karlsruhe, Germany.

compare in Section 3 classes of functions generating the same metric. We especially characterize the maximal generator of an integral probability metric. In Section 4 we give conditions on the generator which induce interesting properties of the probability metric. Special emphasis is given to the relationship between convergence in the metric and weak convergence. In Section 5 we apply these results to several examples, like the Kolmogorov metric, the total variation metric and the stop-loss metric.

2. Preliminaries

We first make some remarks about our notation. Sets of functions are mostly denoted by capital *fraktur* letters such as $\mathfrak{F}, \mathfrak{D}, \mathfrak{R}, \mathfrak{B}, \cdots$, whereas we use *calligraphic* letters such as $\mathcal{A}, \mathcal{B}, \mathcal{S} \cdots$ for σ-algebras. Sets of (signed) measures are denoted by open-face letters such as $\mathbb{M}, \mathbb{P}, \cdots$.

Let $(S, \mathcal{S})$ be a measure space and $b: S \to [1, \infty)$ a measurable function, called a *weight function*. We consider the set $\mathfrak{B}_b$ of measurable functions $f: S \to \mathbb{R}$, for which

$$\|f\|_b := \sup_{s \in S} \frac{|f(s)|}{b(s)} < \infty.$$

For a signed measure μ on $\mathcal{S}$ we denote the positive and negative variation by μ^+ resp. μ^-. As usual $|\mu| := \mu^+ + \mu^-$ is the total variation. Integrals are sometimes written in the functional form $\mu(f) := \int f \, d\mu := \int f \, d\mu^+ - \int f \, d\mu^-$. Notice that $\mu(f)$ exists and is finite if and only if $\mu^+(|f|) + \mu^-(|f|) < \infty$.

The set of all signed measures μ on $\mathcal{S}$ with $|\mu|(b) = \mu^+(b) + \mu^-(b) < \infty$ is denoted by $\mathbb{M}_b$. We write $\mathbb{P}$ for the set of all probability measures (p.m.) on $\mathcal{S}$, and $\mathbb{P}_b := \mathbb{P} \cap \mathbb{M}_b$ is the restriction of $\mathbb{M}_b$ to $\mathbb{P}$. $\mathbb{P}_b$ is non-void as it contains all the p.m. with finite support. $\mathbb{M}_b^N$ is the set of all signed measures with $\mu(S) = 0$. Notice that the difference of two p.m. lies in $\mathbb{M}_b^N$ and that every measure in $\mathbb{M}_b^N$ is a multiple of such a difference, i.e. $\mathbb{M}_b^N$ is the linear span of $\mathbb{P}_b - \mathbb{P}_b$.

For the formulation of our first lemmas we need some notions from functional analysis, which can be found, for example, in Choquet (1969), §22, or Robertson and Robertson (1966).

A pair (E, F) of vector spaces is said to be in *duality* if there is a bilinear mapping $\langle \cdot, \cdot \rangle: E \times F \to \mathbb{R}$. The duality is said to be *strict* if, for each $0 \neq x \in E$, there is a $y \in F$ with $\langle x, y \rangle \neq 0$ and for each $0 \neq y \in F$ there is an $x \in E$ with $\langle x, y \rangle \neq 0$.

Lemma 2.1. *$\mathbb{M}_b$ and $\mathfrak{B}_b$ are in strict duality under the bilinear mapping*

$$\begin{aligned} &\langle \cdot, \cdot \rangle: \mathbb{M}_b \times \mathfrak{B}_b \to \mathbb{R} \\ &\langle \mu, f \rangle := \mu(f). \end{aligned} \tag{2.1}$$

Proof. Evidently $\mathfrak{B}_b$ and $\mathbb{M}_b$ are vector spaces. For $f \in \mathfrak{B}_b$ we have $|f| \leqq \|f\|_b \cdot b$, and hence

$$|\mu(f)| \leqq \mu^+(|f|) + \mu^-(|f|) \leqq \|f\|_b(\mu^+(b) + \mu^-(b)) < \infty$$

for $\mu \in \mathbb{M}_b$. Thus the mapping $\langle \cdot, \cdot \rangle$ is well defined. It remains to show strictness of the duality.

(i) $\mathfrak{B}_b$ contains the indicator functions of all sets $A \in \mathcal{S}$, as $b \geqq 1$. Therefore $\mu(f) = 0$ for all $f \in \mathfrak{B}_b$ implies $\mu(A) = 0$ for all $A \in \mathcal{S}$, and thus $\mu \equiv 0$.

(ii) $\mathbb{M}_b$ contains all one-point measures δ_s, $s \in S$. Hence $\mu(f) = 0$ for all $\mu \in \mathbb{M}_b$ implies $\delta_s(f) = f(s) = 0$ for all $s \in S$ and consequently $f \equiv 0$.

Remark. In part (i) of the proof we needed the requirement $b \geqq 1$ for the weight function. Sometimes there is a naturally given weight function b', which only fulfils $b' \geqq 0$. We can then use $b := b' + 1$, leading to $\mathbb{M}_b = \mathbb{M}_{b'}$ and $\mathfrak{B}_{b'} \subset \mathfrak{B}_b$, i.e. the measure space remains the same and even more functions can be handled.

Unfortunately the duality $(\mathbb{M}_b^N, \mathfrak{B}_b)$ is not strict, as $\mu(f) = 0$ for all $\mu \in \mathbb{M}_b^N$ only implies that f is constant. But strict duality can be obtained by identifying functions that differ only by a constant. Formally, we define an equivalence relation $f \sim g$ if and only if $f - g$ is constant. Denoting the corresponding quotient space by $\mathfrak{B}_{b/\sim}$ we get the following lemma.

Lemma 2.2. *$\mathbb{M}_b^N$ and $\mathfrak{B}_{b/\sim}$ are vector spaces in strict duality under the bilinear mapping* (2.1).

A crucial role in our further investigations is played by the bipolar theorem. The *polar* M^0 of a set $M \subset E$ (in the duality (E, F)) is defined by

$$M^0 := \{y \in F : |\langle x, y \rangle| \leqq 1 \text{ for all } x \in M\}.$$

The following theorem is known as the *bipolar theorem* (see e.g. Robertson and Robertson (1966), p. 35),

Theorem 2.3. (*Bipolar theorem*). *Suppose E and F are in strict duality and $X \subset E$. Then X^{00} is the $\sigma(E, F)$-closure of the absolutely convex hull of X.*

3. Maximal generators

Let $(S, \mathcal{S})$ be an arbitrary measure space and let $b: S \to \mathbb{R}$ be a weight function. A mapping $d: \mathbb{P}_b \times \mathbb{P}_b \to [0, \infty]$ is called a *probability metric* if it possesses the following properties.

(i) $d(P_1, P_2) = 0$ if and only if $P_1 = P_2$.

(ii) $d(P_1, P_2) = d(P_2, P_1)$.

(iii) $d(P_1, P_3) \leqq d(P_1, P_2) + d(P_2, P_3)$.

If (i) is replaced by the weaker requirement $d(P, P) = 0$ for all $P \in \mathbb{P}_b$, then we speak of a *probability semimetric.*

https://doi.org/10.2307/1428011 Published online by Cambridge University Press

In this paper we only consider probability (semi)metrics that are generated by integrals. For $\mathfrak{F} \subset \mathfrak{B}_b$ we define an *integral probability (semi)metric* $d_{\mathfrak{F}}$ on $\mathbb{P}_b$ by

$$d_{\mathfrak{F}}(P, Q) := \sup_{f \in \mathfrak{F}} \left| \int f \, dP - \int f \, dQ \right|. \tag{3.1}$$

Remarks. (1) As is common use in the theory of probability metrics, the distance between two p.m. is allowed to be infinite, see e.g. Rachev (1991), p. 10ff.

(2) In Zolotarev (1983) and Rachev (1991) probability metrics defined as in (3.1) are called *metrics with ζ-structure*.

(3) The metric $d_{\mathfrak{F}}$ is induced by a seminorm $\|\cdot\|_{\mathfrak{F}}$ on $\mathbb{M}_b^N$. If we define

$$\|\mu\|_{\mathfrak{F}} := \sup_{f \in \mathfrak{F}} |\mu(f)|, \tag{3.2}$$

then $d_{\mathfrak{F}}(P, Q) = \|P - Q\|_{\mathfrak{F}}$.

(4) The function $d_{\mathfrak{F}}$ is obviously a probability semimetric. It is a metric if and only if $\mathfrak{F}$ separates points in $\mathbb{M}_b^N$.

Next we compare different classes of functions that generate the same probability metric.

Definition 3.1. *Let $\mathfrak{F} \subset \mathfrak{B}_b$. The set $\mathfrak{R}_{\mathfrak{F}}$ of all functions $f \in \mathfrak{B}_b$ with the property*

$$|P(f) - Q(f)| \leqq d_{\mathfrak{F}}(P, Q) \quad \text{for all} \quad P, Q \in \mathbb{P}_b \tag{3.3}$$

is called a maximal generator.

Remark. Since $\mathbb{M}_b^N$ is the linear span of $\mathbb{P}_b - \mathbb{P}_b$, $f \in \mathfrak{R}_{\mathfrak{F}}$ holds if and only if $|\mu(f)| \leqq \|\mu\|_{\mathfrak{F}}$ for all $\mu \in \mathbb{M}_b^N$.

A direct consequence of the definition is the following result.

Lemma 3.2. *Let $\mathfrak{F} \subset \mathfrak{D} \subset \mathfrak{B}_b$ and $P, Q \in \mathbb{P}_b$.*
(a) *$d_{\mathfrak{F}}(P, Q) \leqq d_{\mathfrak{D}}(P, Q)$.*
(b) *$\mathfrak{R}_{\mathfrak{F}} \subset \mathfrak{R}_{\mathfrak{D}}$.*
(c) *If $\mathfrak{D} \subset \mathfrak{R}_{\mathfrak{F}}$, then $d_{\mathfrak{D}}$ and $d_{\mathfrak{F}}$ are identical.*

The next two results show that $\mathfrak{R}_{\mathfrak{F}}$ is absolutely convex, contains the constant functions and is closed under linear mixtures.

Theorem 3.3. *Let $\mathfrak{F}$ be an arbitrary generator of $d_{\mathfrak{F}}$. Then:*
(a) *$\mathfrak{R}_{\mathfrak{F}}$ contains the convex hull of $\mathfrak{F}$;*
(b) *$f \in \mathfrak{R}_{\mathfrak{F}}$ implies $\alpha f + \beta \in \mathfrak{R}_{\mathfrak{F}}$ for all $\alpha \in [-1, 1]$ and $\beta \in \mathbb{R}$; and*
(c) *if the sequence $(f_n)_{n \in \mathbb{N}} \subset \mathfrak{R}_{\mathfrak{F}}$ converges uniformly to f, then $f \in \mathfrak{R}_{\mathfrak{F}}$.*

Theorem 3.4. *Let $(\Omega, \mathcal{A}, \rho)$ be a probability space and let $f: \Omega \times S \to \mathbb{R}$ be a $\mathcal{A} \otimes \mathcal{S}$-measurable function that fulfils the following assumptions.*
(i) *$f(\omega, \cdot) \in \mathfrak{F}$ for all $\omega \in \Omega$.*

https://doi.org/10.2307/1428011 Published online by Cambridge University Press

(ii) *There exists a ρ-integrable function $c:\Omega\to\mathbb{R}_{\geqq 0}$ with $|f(\omega, s)|\leqq c(\omega)b(s)$ for all $\omega\in\Omega$, $s\in S$.*
Then $g(\cdot):=\int f(\omega,\cdot)\rho(d\omega)$ exists and belongs to $\mathfrak{R}_{\mathfrak{F}}$.

Proof. Since $|f(\omega, x)|\leqq c(\omega)\cdot b(x)$ we have for all $\mu\in\mathbb{M}_b$:

$$\int\int |f(\omega, x)|\,\rho(d\omega)\,|\mu|(dx)\leqq\int c(\omega)\rho(d\omega)\cdot\int b(x)\,|\mu|\,(dx)<\infty. \tag{3.4}$$

Specializing $\mu=\delta_s$, $s\in S$, we can infer the existence of $g(s)=\int f(\omega, s)\rho(d\omega)$. Now (3.4) and (ii) imply $\|g\|_b\leqq\int c\,d\rho<\infty$. Hence $g\in\mathfrak{B}_b$ and we can apply Fubini's theorem. Thus we have for $P, Q\in\mathbb{P}_b$

$$\begin{aligned}|P(g)-Q(g)|&\leqq\int\rho(d\omega)\left|\int P(ds)f(\omega, s)-\int Q(ds)f(\omega, s)\right|\\&\leqq\int\rho(d\omega)d_{\mathfrak{F}}(P, Q)=d_{\mathfrak{F}}(P, Q).\end{aligned}$$

This yields $g\in\mathfrak{R}_{\mathfrak{F}}$.

The next theorem is one of the main results of this paper. It gives a characterization of the maximal generator.

Theorem 3.5. *$\mathfrak{R}_{\mathfrak{F}}$ is the $\sigma(\mathfrak{B}_b, \mathbb{M}_b)$-closure of the absolutely convex hull of $\mathfrak{F}$ and the constant functions.*

Proof. The assertion follows from the bipolar theorem for the duality $(\mathbb{M}_b^N, \mathfrak{B}_{b/\sim})$, if we can show $\mathfrak{R}_{\mathfrak{F}/\sim}=(\mathfrak{F}_{/\sim})^{00}$. But by the definition of $\|\cdot\|_{\mathfrak{F}}$ we have

$$\begin{aligned}f\in(\mathfrak{F}/\sim)^{00}&\Leftrightarrow|\mu(f)|\leqq 1 && \forall\mu\in(\mathfrak{F}/\sim)^0\\&\Leftrightarrow|\mu(f)|\leqq 1 && \forall\mu\in\mathbb{M}_b^N\ \text{ with }\ \|\mu\|_{\mathfrak{F}}\leqq 1\\&\Leftrightarrow|\mu(f)|\leqq\|\mu\|_{\mathfrak{F}} && \forall\mu\in\mathbb{M}_b^N\\&\Leftrightarrow f\in\mathfrak{R}_{\mathfrak{F}/\sim}.\end{aligned}$$

Theorem 3.5 is of rather theoretical nature. As the $\sigma(\mathfrak{B}_b, \mathbb{M}_b)$-topology is hard to handle, it is not very useful for applications. In our next result, however, we give a sufficient condition for $\mathfrak{F}=\mathfrak{R}_{\mathfrak{F}}$, which is very easy to check.

Corollary 3.6. *If $\mathfrak{F}\subset\mathfrak{D}\subset\mathfrak{R}_{\mathfrak{F}}$, and $\mathfrak{D}$ is absolutely convex, contains the constant functions and is closed with respect to pointwise convergence, then $\mathfrak{D}=\mathfrak{R}_{\mathfrak{F}}$.*

Proof. It is sufficient to show that $\mathfrak{D}$ is closed with respect to the topology $\sigma(\mathfrak{B}_b, \mathbb{M}_b)$. Since $\mathbb{M}_b$ includes all one-point measures, the $\sigma(\mathfrak{B}_b, \mathbb{M}_b)$-topology is finer than the topology of pointwise convergence. Hence each set, which is closed under pointwise convergence, is also closed with respect to $\sigma(\mathfrak{B}_b, \mathbb{M}_b)$.

https://doi.org/10.2307/1428011 Published online by Cambridge University Press

4. Convergence and uniformity

There is a particular interest in probability metrics that metrize weak convergence. Therefore, we now investigate the relationship between structural properties of $\mathfrak{F}$ and weak convergence. From now on we assume that S is a Polish space, i.e. the topological space $(S, \mathfrak{T})$ is metrizable by some metric d such that (S, d) is a complete separable metric space.

Definition 4.1. *Let S be a Polish space and let $b: S \to [1, \infty)$ be any weight function. Let $d_{\mathfrak{F}}$ be some probability metric on $\mathbb{P}_b$. Then $d_{\mathfrak{F}}$ has*:

(a) *property* (W_1), *if $d_{\mathfrak{F}}$ metrizes weak convergence*;

(b) *property* (W_2), *if* $\lim_{n\to\infty} d_{\mathfrak{F}}(P_n, Q_n) = d_{\mathfrak{F}}(P, Q)$ *for all weak convergent sequences* (P_n), $(Q_n) \subset \mathbb{P}_b$ *with limits* $P, Q \in \mathbb{P}_b$;

(c) *property* (W_3), *if* $\liminf_{n\to\infty} d_{\mathfrak{F}}(P_n, Q_n) \geqq d_{\mathfrak{F}}(P, Q)$ *for all weak convergent sequences* (P_n), $(Q_n) \subset \mathbb{P}_b$ *with limits* $P, Q \in \mathbb{P}_b$.

Remarks: (1) The implications $(W_1) \Rightarrow (W_2) \Rightarrow (W_3)$ are obvious, but in the sequel we will see that none of them can be reversed.

(2) Property (W_2) is equivalent to the following condition:

$$(W_2^*) \quad \text{if } (P_n) \text{ converges weakly to } P, \text{ then } d_{\mathfrak{F}}(P_n, P) \to 0.$$

If $d_{\mathfrak{F}}$ has this property, then $\mathfrak{F}$ is sometimes called *a uniform class with respect to weak convergence*, see Rachev (1991), p. 75.

In the following theorem we denote by $\mathfrak{C}_b$ the set of all bounded continuous functions.

Theorem 4.2. *If* $\mathfrak{F} \subset \mathfrak{C}_b$, *then* (W_3) *holds.*

Proof. Let $a := \liminf_{n\to\infty} d_{\mathfrak{F}}(P_n, Q_n)$ and $\varepsilon > 0$. Then there is a subsequence $(k_n) \subseteq \mathbb{N}$ and a $n_0 \in \mathbb{N}$ such that for all $n \geqq n_0$ and all $f \in \mathfrak{F}$ we have:

$$|P_{k_n}(f) - Q_{k_n}(f)| \leqq a + \varepsilon.$$

Hence, if (P_n) and (Q_n) are weak convergent sequences with limits P, Q and $\mathfrak{F} \subset \mathfrak{C}_b$ then $|P(f) - Q(f)| \leqq a + \varepsilon$ for all $f \in \mathfrak{F}$. But this implies

$$d_{\mathfrak{F}}(P, Q) = \sup_{f \in \mathfrak{F}} |P(f) - Q(f)| \leqq a + \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, the assertion follows.

For an arbitrary function f we define the *span* of f by $\mathrm{sp}(f) := \sup f - \inf f$. A set $\mathfrak{F}$ of functions is said to have *uniformly bounded span*, if

$$\sup_{f \in \mathfrak{F}} \mathrm{sp}\,(f) < \infty.$$

https://doi.org/10.2307/1428011 Published online by Cambridge University Press

The following theorem can be found in Bhattacharya and Ranga Rao (1976), p. 16.

Theorem 4.3. *An integral probability metric $d_{\mathfrak{F}}$ has property* (W_2) *if and only if* $\mathfrak{F}$ *is equicontinuous and has uniformly bounded span.*

Corollary 4.4. *Property* (W_2) *holds if and only if the function* $d^*(x, y) := d_{\mathfrak{F}}(\delta_x, \delta_y)$ *is continuous and bounded.*

A necessary and sufficient condition on $\mathfrak{F}$ for (W_1) to hold is, to our knowledge, unknown. From the preceding theorem it is evident that it is necessary for $\mathfrak{F}$ to have uniformly bounded span and to be equicontinuous. Another necessary condition is the following: the semimetric d^* defined in Corollary 4.4 is topologically equivalent to d.

Example. Let $S = \mathbb{R}$ and let

$$\mathfrak{F} := \{f : \mathbb{R} \to \mathbb{R} : \|f\|_L \leqq 1,\ \|f\|_\infty \leqq 1,\ |f(x) - f(0)| \leqq |1/x|,\ x \neq 0\},$$

where $\|\cdot\|_L$ is the so called *Lipschitz-norm* defined on an arbitrary metric space (S, d) as

$$\|f\|_L := \sup_{x \neq y \in S} \frac{|f(x) - f(y)|}{d(x, y)}.$$

Then $d_{\mathfrak{F}}$ has property (W_2), as $\mathfrak{F}$ is uniformly bounded and equicontinuous. But (W_1) does not hold, since $d_{\mathfrak{F}}(\delta_n, \delta_0) = 1/n \to 0$ for $n \to \infty$.

The following sufficient condition for (W_1) can easily be deduced from Theorem 4.3.2 in Rachev (1991).

Theorem 4.5. *If* $\mathfrak{F}$ *has uniformly bounded span, is equicontinuous and contains for every closed set* $A \subset S$ *and all* $n \in \mathbb{N}$ *the function*

$$s \to f_{n,A}(s) := \max\{0, 1/n - d(s, A)\},$$

then $d_{\mathfrak{F}}$ *has the property* (W_1).

A well known example for an integral probability metric that metrizes weak convergence is the *Dudley metric* β. It is generated by the set

$$\mathfrak{F} := \{f : \|f\|_\infty \leqq 1,\ \|f\|_L \leqq 1\}.$$

This metric obviously fulfils the conditions of Theorem 4.5.

There are some more interesting properties of probability metrics. Some of them are most easily defined in terms of random variables. Therefore we sometimes use the notation $d_{\mathfrak{F}}(X, Y) := d_{\mathfrak{F}}(P_X, P_Y)$.

Definition 4.6. *Let* (S, d) *be a metric vector space,* $b : S \to [1, \infty)$ *a weight function and* $d_{\mathfrak{F}}$ *any* (*semi*)*metric on* $\mathbb{P}_b$. *Then* $d_{\mathfrak{F}}$ *is said to have*

(a) *property* (R), *if* $d_{\mathfrak{F}}(\delta_a, \delta_b) = d(a, b)$;
(b) *property* (M), *if* $d_{\mathfrak{F}}(aX, aY) = a d_{\mathfrak{F}}(X, Y)$;
(c) *property* (C), *if* $d_{\mathfrak{F}}(P_1 * Q, P_2 * Q) \leqq d_{\mathfrak{F}}(P_1, P_2)$ *for all p.m. Q*.

Theorem 4.7. *Let* $\mathfrak{R}_{\mathfrak{F}}$ *be the maximal generator of the integral probability metric* $d_{\mathfrak{F}}$ *on* $\mathbb{P}_b$.

(*a*) *Property* (R) *holds if and only if*

$$\sup_{f \in \mathfrak{F}} \frac{|f(x) - f(y)|}{d(x, y)} = 1 \quad \text{for all} \quad x, y \in S, x \neq y.$$

(*b*) *Property* (C) *holds if and only if* $\mathfrak{R}_{\mathfrak{F}}$ *is invariant under translations.*

Proof. (a) is trivial.

(b) (i) If $\mathfrak{R}_{\mathfrak{F}}$ is invariant under translations, then $f \in \mathfrak{R}_{\mathfrak{F}}$ implies $f(\cdot + y) \in \mathfrak{R}_{\mathfrak{F}}$ for all $y \in S$. Hence

$$\begin{aligned} d_{\mathfrak{F}}(P_1 * Q, P_2 * Q) &= \sup_{f \in \mathfrak{R}_{\mathfrak{F}}} \left| \int Q(dy) P_1(f(\cdot + y)) - \int Q(dy) P_2(f(\cdot + y)) \right| \\ &\leqq \sup_{f \in \mathfrak{R}_{\mathfrak{F}}} \int Q(dy)\, d_{\mathfrak{F}}(P_1, P_2) = d_{\mathfrak{F}}(P_1, P_2). \end{aligned}$$

Hence (C) holds.

(ii) Now assume that (C) holds. Let $f \in \mathfrak{R}_{\mathfrak{F}}$ and $P_1, P_2 \in \mathbb{P}_b$. Define $Q := \delta_y$, $y \in S$. Then we can infer

$$\begin{aligned} |P_1(f(\cdot + y)) - P_2(f(\cdot + y))| &= |P_1 * Q(f) - P_2 * Q(f)| \\ &\leqq d_{\mathfrak{F}}(P_1 * Q, P_2 * Q) \overset{(C)}{\leqq} d_{\mathfrak{F}}(P_1, P_2). \end{aligned}$$

Hence $f(\cdot + y) \in \mathfrak{R}_{\mathfrak{F}}$.

5. Examples

5.1. *The Kolmogorov metric.* A well known probability metric on $S = \mathbb{R}$ is the *Kolmogorov metric* ρ defined by

$$\rho(X, Y) := \sup_{t \in \mathbb{R}} |F_X(t) - F_Y(t)|.$$

Since $F_X(t) = \int \mathbf{1}_{[t,\infty)}\, dP_X$, the Kolmogorov metric is an integral probability metric generated by the set $\mathfrak{F}_\rho$ of all functions $\mathbf{1}_{[t,\infty)}$, $t \in \mathbb{R}$. One can use $b(s) \equiv 1$ as weight function, so that $\mathbb{P}_b$ consists of all probability measures on $\mathbb{R}$.

The maximal generator of ρ can be characterized in terms of total variation. We

https://doi.org/10.2307/1428011 Published online by Cambridge University Press

denote the set of all functions of bounded variation by BV $(\mathbb{R})$. For a partition $Z = [x_0, x_1, \cdots, x_n]$ of $\mathbb{R}$ with $x_0 < x_1 < \cdots < x_n$ we define

$$V(f, Z) := \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|.$$

Then the total variation of a function $f \in \mathrm{BV}(\mathbb{R})$ is defined by

$$V(f) := \sup_Z V(f, Z).$$

Lemma 5.1. *The set* $\mathfrak{F}_1 := \{f \in \mathrm{BV}(\mathbb{R}) : V(f) \leqq 1\}$ *is closed with respect to pointwise convergence.*

Proof. Endow BV $(\mathbb{R})$ with the topology of pointwise convergence. Then for a fixed partition Z the functional $f \to V(f, Z)$ is continuous. Hence the functional $f \to V(f) = \sup_Z V(f, Z)$ is lower semicontinuous, as it is a supremum of continuous functionals. Thus the level set $\{V(f) \leqq 1\}$ is closed.

Theorem 5.2. *The maximal generator* $\mathfrak{R}_{\mathfrak{F}}$ *of the Kolmogorov metric* ρ *is the set of all functions* $f \in \mathrm{BV}(\mathbb{R})$ *with total variation* $V(f) \leqq 1$.

Proof. Let $\mathfrak{F}_1$ be the set of all functions $f \in \mathrm{BV}(\mathbb{R})$ with $V(f) \leqq 1$. Then obviously $\mathfrak{F}_\rho \subset \mathfrak{F}_1$. The convex hull of $\mathfrak{F}_\rho$ is the set of all increasing step functions that assume values in $[0, 1]$. But every increasing function with range in $[0, 1]$ can be approximated uniformly by such step functions. Thus Theorem 3.3 implies that $\mathfrak{R}_{\mathfrak{F}}$ contains all monotone functions f with $V(f) \leqq 1$. Now the decomposition theorem of Jordan tells us that every function f with $V(f) \leqq 1$ can be written as a convex combination of two monotone functions with this property. Hence applying Theorem 3.3 once more yields $\mathfrak{F}_1 \subset \mathfrak{R}_{\mathfrak{F}}$. The assertion now follows from Lemma 5.1 and Corollary 3.6.

Remark. Rachev (1991) claims on p.73 that the Kolmogorov metric is generated by the set of all (Lebesgue) a.e. differentiable functions f with $\int |f'(x)|\, dx \leqq 1$. This statement is not true since this set of functions contains all step functions.

Theorem 5.3. *The Kolmogorov metric* ρ *has the properties* (C) *and* (W_3).

Proof. (a) The maximal generator of ρ obviously is invariant under translations. Hence Theorem 4.7 implies (C).

(b) By Theorem 4.2, the existence of a generator $\mathfrak{D} \subset \mathfrak{C}_b$ is sufficient for (W_3). Such a generator is given by the set of all functions

$$f_{a,b}(x) = \begin{cases} 0 & \text{for } x < a, \\ (x-a)/(b-a) & \text{for } a \leqq x \leqq b, \\ 1 & \text{for } x > b, \end{cases}$$

$a, b \in \mathbb{R}$, $a < b$.

5.2. *The total variation metric.* The set of all signed measures on an arbitrary measure space $(S, \mathcal{S})$ can be endowed with the so-called *total variation norm* $\|\mu\| := |\mu|(S)$. The corresponding *total variation metric* σ on the set of all probability measures is then defined by $\sigma(P, Q) := |P - Q|(S)$, see Zolotarev (1983).

This is an integral probability metric. Choose $b(s) \equiv 1$ and define $\mathcal{F}_\sigma := \{2 \cdot 1_B : B \in \mathcal{S}\}$. Then $\mathcal{F}_\sigma$ is a generator of σ, as

$$\|\mu\| = 2 \cdot \sup_{A \in \mathcal{S}} |\mu(A)|$$

for all $\mu \in \mathbb{M}_b^N$.

The proof of the following theorem is similar to that of Theorem 5.2.

Theorem 5.4. *The maximal generator* $\mathfrak{R}_\mathfrak{F}$ *of the total variation metric* σ *is the set of all measurable functions* $f : S \to \mathbb{R}$ *with* $\mathrm{sp}(f) \leqq 2$.

Theorem 5.5. *The total variation metric* σ *has the properties* (C) *and* (W_3).

Proof. (a) Property (C) follows immediately from Theorem 4.7.

(b) We claim that $\mathfrak{D} := \mathfrak{R}_\mathfrak{F} \cap \mathfrak{C}_b$ is a generator of σ. Property (W_3) then follows from Theorem 4.2. Let $f \in \mathfrak{R}_\mathfrak{F}$ and $P, Q \in \mathbb{P}_b$. Define $\mu := P + Q$. It is well known that the continuous functions are dense in $\mathcal{L}_1(\mu)$, see e.g. Hewitt and Stromberg (1965), Theorem 13.21. Thus there is a sequence $(\phi_n) \subset \mathfrak{D}$ with $\int |f - \phi_n| \, d\mu \to 0$. This yields $|P(f) - Q(f)| \leqq \mu(|f - \phi_n|) + d_\mathfrak{D}(P, Q) \to d_\mathfrak{D}(P, Q)$. Hence $f \in \mathfrak{R}_\mathfrak{D}$. Thus we have $\mathfrak{D} \subset \mathfrak{R}_\mathfrak{F} \subset \mathfrak{R}_\mathfrak{D}$ and therefore $d_\mathfrak{D} = \sigma$.

5.3. *The stop–loss metric.* Motivated by risk-theoretical considerations, Rachev and Rüschendorf (1990) defined and investigated several so called *stop–loss metrics.* The most important one is

$$d_{\mathrm{sl}}(X, Y) := \sup_{t \in R} |E(X - t)^+ - E(Y - t)^+|,$$

defined for random variables with finite expectation. This metric was already defined in Gerber (1981).

If we define on $S = \mathbb{R}$ the weight function $b(s) := 1 + |s|$, then d_{sl} is an integral probability metric on $\mathbb{P}_b$ that is generated by the set $\mathfrak{F}_{\mathrm{sl}}$ of functions $s \to \phi_t(s) = (s - t)^+$, $t \in \mathbb{R}$.

Before we can characterize the corresponding maximal generator, we need some facts pertaining to differences of convex functions. We follow the notation of Roberts and Varberg (1973). Let I be a closed interval with endpoints a, b. For a partition $Z = \{x_0, x_1, \cdots, x_n\} \subset I$ with $x_0 < x_1 < \cdots < x_n$ and a function $f : I \to \mathbb{R}$ we define

$$\Box_i f := \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} \quad \text{and} \quad K(f, Z) := \sum_{i=1}^{n-1} |\Box_{i+1} f - \Box_i f|.$$

Further, $K_a^b(f) := \sup K(f, Z)$, where the supremum is taken over all partitions Z. In the case $I = \mathbb{R}$ we write $K(f)$ for short. It is clear that

$$K(f) := \lim_{\substack{a \to -\infty \\ b \to \infty}} K_a^b(f). \tag{5.1}$$

For the case where $[a, b]$ is a compact interval, a thorough treatment of the functional K_a^b can be found in Section 14 of Roberts and Varberg (1973). Using (5.1), these results can easily be carried over to the case $I = \mathbb{R}$, yielding the following results.

Lemma 5.6. (a) $K(f) < \infty$, *if and only if* f *is the difference of two Lipschitz-continuous convex functions.*

(b) *If* $K(f) < \infty$, *then the left and right derivatives* $D^-f(x)$ *and* $D^+f(x)$ *exist for all* $x \in \mathbb{R}$ *and* $K(f) = V(D^-f) = V(D^+f)$.

By Lemma 5.6 (b) D^+f is of bounded variation if $K(f) < \infty$. Hence $\lim_{t \to -\infty} D^+f(t)$ exists. Thus we can define a functional K^* by

$$K^*(f) := K(f) + \left| \lim_{t \to -\infty} D^+f(t) \right|.$$

This functional can alternatively be defined as follows. For a partition Z define

$$K^*(f, Z) := |\Box_1 f| + \sum_{i=1}^{n-1} |\Box_{i+1} f - \Box_i f|.$$

Then $K^*(f) = \sup_Z K^*(f, Z)$.

Using this characterization, the following lemma can be proved similarly to Lemma 5.1.

Lemma 5.7. *The set* $\mathfrak{K}_1 := \{f : \mathbb{R} \to \mathbb{R} : K^*(f) \leqq 1\}$ *is closed with respect to pointwise convergence.*

Theorem 5.8. *The maximal generator* $\mathfrak{R}_{\mathfrak{F}}$ *of the stop-loss metric* d_{sl} *is the set of all functions* $f : \mathbb{R} \to \mathbb{R}$ *with* $K^*(f) \leqq 1$.

Proof. (a) Let $\mathfrak{K}_1 := \{f : \mathbb{R} \to \mathbb{R} : K^*(f) \leqq 1\}$. For $\phi_t(s) := (s - t)^+$, $t \in \mathbb{R}$, we have $D^+\phi_t = \mathbf{1}_{[t,\infty)}$, and therefore Lemma 5.6 implies $K^*(\phi_t) = K(\phi_t) = V(D^+\phi_t) = 1$. Hence $\mathfrak{F}_{\mathrm{sl}} = \{\phi_t : t \in \mathbb{R}\} \subset \mathfrak{K}_1$.

(b) Next we show $\mathfrak{K}_1 \subset \mathfrak{R}_{\mathfrak{F}}$. From

$$|EX - EY| = \lim_{t \to -\infty} |E(X - t)^+ - E(Y - t)^+| \leqq d_{\mathrm{sl}}(X, Y),$$

we infer $id \in \mathfrak{R}_{\mathfrak{F}}$. Now fix $f \in \mathfrak{K}_1$ and define $\alpha := \lim_{x \to -\infty} D^+f(x)$ and $\beta := K(f)$. Then $f(x) =: \alpha x + \beta f_1(x)$, where f_1 has the following properties.

(I) $K(f_1) \leqq 1$.

(II) $|\lim_{x\to-\infty} D^+ f_1(x)| = 0$.

Hence, if we can show that every function that fulfils (I) and (II) is contained in $\mathfrak{R}_{\mathfrak{F}}$, then $id \in \mathfrak{R}_{\mathfrak{F}}$ implies $\mathfrak{K}_1 \subset \mathfrak{R}_{\mathfrak{F}}$.

Therefore suppose that f fulfils (I) and (II), and let $g := D^+ f$. Since $V(g) = K(f) \leqq 1$, the function g has the following properties.

(I′) $V(g) \leqq 1$.

(II′) $\lim_{x\to-\infty} g(x) = 0$.

From Jordan's decomposition theorem, we can deduce that there is a $\gamma \in [0, 1]$, such that g can be written as $g = \gamma g_1 - (1-\gamma) g_2$, where g_1, g_2 are increasing functions that fulfil (I′) and (II′) as well. Thus we can assume without loss of generality that g is increasing. Hence g can be approximated monotonely by increasing step functions

$$h_n(x) := \sum_{i=1}^{n} \alpha_{in} \cdot \mathbf{1}_{[\beta_{in},\infty)} \qquad \text{with } \alpha_{in} \geqq 0, \sum_{i=1}^{n} \alpha_{in} \leqq 1, \beta_{in} \in \mathbb{R}.$$

Hence, by Theorem 3.3

$$f_n(x) := f(0) + \int_0^x h_n(t)\,dt = \sum_{i=1}^{n} \alpha_{in} \cdot \phi_{\beta_{in}}(x) + \text{constant} \tag{5.2}$$

is contained in $\mathfrak{R}_{\mathfrak{F}}$, and the monotone convergence theorem implies that (f_n) converges to f, from above on $(-\infty, 0)$ and from below on $[0, \infty)$. Applying the monotone convergence theorem once more, we get

$$\begin{aligned}\lim_{n\to\infty} P(f_n) &= \lim_{n\to\infty} (P(f_n \cdot \mathbf{1}_{(-\infty,0)}) + P(f_n \cdot \mathbf{1}_{[0,\infty)}))\\ &= P(f \cdot \mathbf{1}_{(-\infty,0)}) + P(f \cdot \mathbf{1}_{[0,\infty)}) = P(f)\end{aligned}$$

for every $P \in \mathbb{P}_b$. Since $f_n \in \mathfrak{R}_{\mathfrak{F}}$, this implies for arbitrary $P, Q \in \mathbb{P}_b$

$$|P(f) - Q(f)| = \lim_{n\to\infty} |P(f_n) - Q(f_n)| \leqq d_{\text{sl}}(P, Q).$$

Hence $f \in \mathfrak{R}_{\mathfrak{F}}$ and thus we have shown $\mathfrak{K}_1 \subset \mathfrak{R}_{\mathfrak{F}}$.

(c) By Lemma 5.7, $\mathfrak{K}_1$ is closed with respect to pointwise convergence. It is easy to see that $\mathfrak{K}_1$ is absolutely convex and contains the constant functions. Thus Theorem 3.6 implies $\mathfrak{K}_1 = \mathfrak{R}_{\mathfrak{F}}$.

Remark. Rachev and Rüschendorf (1990) define a probability metric θ_1 generated by

$$\mathfrak{F}_1 := \left\{ f : \mathbb{R} \to \mathbb{R} : f'' \text{ exists and } \int |f''(x)|\,dx \leqq 1 \right\}.$$

https://doi.org/10.2307/1428011 Published online by Cambridge University Press

They show that $\theta_1(X, Y) = d_{sl}(X, Y)$ if $EX = EY$. But, for $EX \neq EY$, $\theta_1(X, Y)$ is not finite, as $\mathfrak{F}_1$ contains all functions $s \to \alpha s$, $\alpha \in \mathbb{R}$. Thus $\mathfrak{F}_1$ is not a generator of d_{sl}. But if we modify $\mathfrak{F}_1$ to

$$\mathfrak{F}_1^* := \left\{ f : \mathbb{R} \to \mathbb{R} : f'' \text{ exists and } \left| \lim_{x \to -\infty} f'(x) \right| + \int |f''(x)|\, dx \leqq 1 \right\},$$

we get a generator of d_{sl}. To see this you only have to observe that we have $K(f) = \int |f''(x)|\, dx$ for any twice differentiable function f, cf. Roberts and Varberg (1973), p. 28, problem D (3).

Theorem 5.9. *The probability metric d_{sl} has the properties* (R) *and* (C), *but none of the properties* (W_1)–(W_3).

Proof. (a) The properties (R) and (C) follow immediately from Theorem 4.7. For (C) notice that the functional K^* is invariant under translations.

(b) We give the following counterexample for (W_3). Let

$$P_n := \frac{n-1}{n} \delta_0 + \frac{1}{n} \delta_n \xrightarrow{w} \delta_0 =: P,$$

and

$$Q_n := \frac{n-1}{n} \delta_1 + \frac{1}{n} \delta_{n/2} \xrightarrow{w} \delta_1 =: Q.$$

Then $d_{sl}(P_n, Q_n) = \frac{1}{2}$, $n \in \mathbb{N}$, but $d_{sl}(P, Q) = 1$. Hence (W_1) – (W_3) cannot hold.

If G is a non-negative unbounded continuous function, then the joint convergence

$$P_n \xrightarrow{w} P \qquad \text{and} \qquad \int G\, dP_n \to \int G\, dP$$

is called *G-weak convergence*, see Rachev (1991), Def. 4.2.2. Using this notion, the following weakening of (W_3) can be proved for d_{sl}.

Theorem 5.10. *Define $G(s) = s^+$ and let (P_n), (Q_n) be G-weak convergent sequences with limits P and Q. Then*

$$\liminf_{n \to \infty} d_{sl}(P_n, Q_n) \geqq d_{sl}(P, Q). \tag{5.3}$$

Proof. The functions $f_t(x) := \phi_t(x) - G(x) = (x - t)^+ - x^+$, $t \in \mathbb{R}$ are bounded and continuous. Therefore G-weak convergence of (P_n) to P implies $\int \phi_t\, dP_n \to \int \phi_t\, dP$. Hence (5.3) can be proved similarly to Theorem 4.2.

The following example shows that $\mathfrak{F}_{sl}$ is not a uniform class with respect to G-weak convergence. Let

$$P_n := \frac{n-1}{n} \delta_0 + \frac{1}{n} \delta_{-n}.$$

Then the sequence (P_n) is G-weak convergent to $P := \delta_0$, but $d_{sl}(P_n, P) = n$. Hence d_{sl} does not metrize G-weak convergence.

5.4. *Further examples.* Another well known integral probability metric is the *Kantorovich metric* ζ_1, which is generated by the set $\mathfrak{L}_1$ of Lipschitz functions f with $\|f\|_L \leqq 1$, see Zolotarev (1983), p. 284 or Dudley (1989), p. 330. It is well known that for $S = \mathbb{R}$

$$\zeta_1(X, Y) = \ell_1(X, Y) := \int |F_X(t) - F_Y(t)|\, dt,$$

see e.g. Rachev (1991), p.6. It is easy to see that $\mathfrak{L}_1$ is the maximal generator of ζ_1 and that ζ_1 has the properties (R), (M), (C) and (W_3).

We have already mentioned the Dudley metric as an integral probability metric that metrizes weak convergence. The most familiar probability metrics with this property are the *Levy metric L* and the *Prohorov metric* π, see Rachev (1991). These two metrics are *not* generated by integrals. This follows from the fact that they both fulfil

$$d(\delta_0, (1-\alpha)\delta_0 + \alpha\delta_{\frac{1}{2}}) = \min\{\alpha, \tfrac{1}{2}\} \tag{5.4}$$

for all $\alpha \in (0, 1)$. But this is not possible for an integral probability metric, since for an arbitrary generator $\mathfrak{F}$

$$d_{\mathfrak{F}}(\delta_0, (1-\alpha)\delta_0 + \alpha\delta_{\frac{1}{2}}) = \|\alpha(\delta_0 - \delta_{\frac{1}{2}})\|_{\mathfrak{F}} = \alpha\, \|\delta_0 - \delta_{\frac{1}{2}}\|_{\mathfrak{F}},$$

in contradiction to (5.4).

Acknowledgements

The content of this paper appears in part in my doctoral thesis, which I have written under the supervision of K. Hinderer. I owe him a debt of gratitude for his expert guidance and encouragement. Moreover I want to thank an anonymous referee for several useful hints.

References

Bhattacharya, R. N. and Ranga Rao, R. (1976) *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

Choquet, G. (1969) *Lectures on Analysis II.* Benjamin, New York.

Dudley, R. M. (1989) *Real Analysis and Probability.* Wadsworth and Brooks, Belmont, CA.

Gerber, H. U. (1981) *An Introduction to Mathematical Risk Theory.* Huebner Foundation Monograph.

Hewitt, E. and Stromberg, K. (1965) *Real and Abstract Analysis.* Springer, Berlin.

Rachev, S. T. (1991) *Probability Metrics and the Stability of Stochastic Models.* Wiley, New York.

Rachev, S. T. and Rüschendorf, L. (1990) Approximation of sums by compound Poisson distributions with respect to stop–loss distances. *Adv. Appl. Prob.* **22,** 350–374.

Roberts, A. W. and Varberg, D. E. (1973) *Convex Functions.* Academic Press, New York.

ROBERTSON, A. P. AND ROBERTSON, W. (1966) *Topological Vector Spaces.* Cambridge University Press, Cambridge.

ZOLOTAREV, V. M. (1983) Probability metrics. *Theory Prob. Appl.* **28,** 278–302.