Randomness criteria in terms of α-divergences

Akio Fujiwara*
Department of Statistical Science, University College London
Gower Street, London WC1E 6BT, United Kingdom.

Abstract

Vovk's randomness criterion characterizes sequences that are random relative to two distinct computable probability measures. The uniqueness of the criterion lies in the fact that, unlike the standard criterion based on the likelihood ratio test, it is expressed in terms of a geometrical quantity, the Hellinger distance, on the space of probability measures. In this paper, we generalize the randomness criterion to a wider class of geometrical quantities, the α-divergences with $-1 < \alpha < 1$. The non-extendibility of the criterion across the boundaries $\alpha = \pm 1$ is investigated in connection with the likelihood ratio test and information geometry.

Keywords and phrases. α-divergence, constructive support, Hellinger distance, information geometry, Kakutani dichotomy, Kolmogorov complexity, Kullback-Leibler divergence, Martin-Löf randomness, ∇^e-geodesic.

1 Introduction

Let $\mathcal{A}$ be a finite set, and let $\mathcal{A}^n$, $\mathcal{A}^*$, and $\mathcal{A}^\infty$ denote the sets of sequences, each comprising elements in $\mathcal{A}$, of length n, of finite length, and of infinite length (one-sided). We denote the null sequence by λ. Further, let $\mathrm{R}_{\mathrm{ML}}(P)$ be the set of Martin-Löf random infinite sequences relative to a computable probability measure P on $\mathcal{A}^\infty$ [1, 2]. The set $\mathrm{R}_{\mathrm{ML}}(P)$ is sometimes called the *constructive support* of P, and is closely related to the measure theoretic notions of singularity and absolute continuity. In fact, given computable probability measures P and Q, they are mutually singular (denoted by $P \perp Q$) if and only if $\mathrm{R}_{\mathrm{ML}}(P) \cap \mathrm{R}_{\mathrm{ML}}(Q) = \emptyset$ [3]. On the other hand, P is absolutely continuous with respect to Q (denoted by $P \ll Q$) if $\mathrm{R}_{\mathrm{ML}}(P) \subset \mathrm{R}_{\mathrm{ML}}(Q)$. The converse implication does not hold in general: a counterexample was given by An. Muchnik [4]. (See also [5], and Section 3.3 below.) This illustrates a delicate aspect of the relationship between $\mathrm{R}_{\mathrm{ML}}(P)$ and $\mathrm{R}_{\mathrm{ML}}(Q)$.

Elements of $\mathrm{R}_{\mathrm{ML}}(P) \cap \mathrm{R}_{\mathrm{ML}}(Q)$ for computable probability measures P and Q are usually characterized by the likelihood ratio test as [5, 6]

$$\mathrm{R}_{\mathrm{ML}}(P) \cap \mathrm{R}_{\mathrm{ML}}(Q) = \left\{ \omega \in \mathrm{R}_{\mathrm{ML}}(P) \;\middle|\; \lim_{n\to\infty} \frac{Q(\omega^n)}{P(\omega^n)} \text{ converges to a positive number} \right\}. \tag{1}$$

*Permanent address: Department of Mathematics, Osaka University, Toyonaka, Osaka 560-0043, Japan (e-mail: fujiwara@math.sci.osaka-u.ac.jp).

Here ω^n is the prefix of ω of length n, and $P(\omega^n) := P(\Gamma_{\omega^n})$, with Γ_a the cylinder set specified by the prefix $a \in \mathcal{A}^*$. The criterion (1) is proved by the effective version of the martingale convergence theorem.

When $\mathrm{R_{ML}}(P) \cap \mathrm{R_{ML}}(Q) \neq \emptyset$, it is expected that the measures P and Q are in some sense "close" to each other. However, the likelihood ratio criterion (1) is not directly connected with the topological structure of the probability space, since it hinges on each individual sequence ω. Vovk [7], on the other hand, derived a rather different criterion by using the Hellinger distance

$$D^{(0)}(p\|q) := 2\sum_{x\in\mathcal{A}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 \tag{2}$$

between two probability measures on $\mathcal{A}$ as follows.

Theorem 1. *Let P and Q be computable probability measures and let $\omega \in \mathrm{R_{ML}}(P)$. Then $\omega \in \mathrm{R_{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and*

$$\sum_{i=1}^{\infty} D^{(0)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) < \infty.$$

Vovk's criterion tells us that for $\omega \in \mathrm{R_{ML}}(P) \cap \mathrm{R_{ML}}(Q)$, the conditional probabilities $P(\,\cdot\,|\omega^{i-1})$ and $Q(\,\cdot\,|\omega^{i-1})$ approach each other sufficiently quickly as measured by the Hellinger distance. In particular, when P and Q are both product measures, say $P = \prod_i P_i$ and $Q = \prod_i Q_i$, the corresponding criterion $\sum_{i=1}^{\infty} D^{(0)}(P_i\|Q_i) < \infty$, which makes no reference to a sequence $\omega \in \mathrm{R_{ML}}(P)$, is nothing but the celebrated Kakutani criterion [8, 9, 10], and we obtain the effective version of the Kakutani dichotomy as follows.

Theorem 2. *When P and Q are computable product measures satisfying $P \overset{loc}{\ll} Q$ (i.e., $P_i \ll Q_i$ for all i), then either $P \ll Q$ or $P \perp Q$ holds, and*
(i) $P \ll Q \iff \mathrm{R_{ML}}(P) \subset \mathrm{R_{ML}}(Q)$,
(ii) $P \perp Q \iff \mathrm{R_{ML}}(P) \cap \mathrm{R_{ML}}(Q) = \emptyset$.

Proof. We need only prove the $\Rightarrow$ of (i). Suppose $\omega \in \mathrm{R_{ML}}(P)$. Then it holds that $P(\omega^n) \neq 0$ for all n, so that $Q(\omega^n) \neq 0$ for all n because of the assumption of local absolute continuity. Moreover, under this assumption, $P \ll Q$ is equivalent to $\sum_{i=1}^{\infty} D^{(0)}(P_i\|Q_i) < \infty$. It then follows from Theorem 1 that $\omega \in \mathrm{R_{ML}}(Q)$. □

Theorem 1 can be regarded as the emergence of the intrinsic topological structure of randomness. In view of information geometry [11], the Hellinger distance is a special example of α-divergence: for $\alpha \neq \pm 1$

$$D^{(\alpha)}(p\|q) := \frac{4}{1-\alpha^2}\left[1 - \sum_{x\in\mathcal{A}} p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}}\right], \tag{3}$$

and $D^{(\pm1)}(p\|q) := \lim_{\alpha\to\pm1} D^{(\alpha)}(p\|q)$. In particular, the (-1)-divergence

$$D^{(-1)}(p\|q) := \sum_{x\in\mathcal{A}} p(x) \log\frac{p(x)}{q(x)}, \tag{4}$$

which is also called the Kullback-Leibler divergence, plays a crucial role in statistical hypothesis testing. Since the notion of randomness is closely related to hypothesis testing, it is natural to ask if Theorem 1 can be extended to other α-divergences. The answer is given by the following

Theorem 3. *Let P and Q be computable probability measures and let $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$. Then $\omega \in \mathrm{R}_{\mathrm{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and*

$$\sum_{i=1}^{\infty} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) < \infty$$

for some (in fact, any) $\alpha \in (-1, 1)$.

It is important to notice that the criterion does not extend to $\alpha = \pm 1$ and beyond. In this sense, Theorem 3 characterizes all the possible criteria for randomness along this line.

This paper is organized as follows. In Section 2, we prove Theorem 3 by modifying Vovk's ingenious argument [7]. In Section 3, we give some examples to get a better perspective on Theorem 3. In particular, we demonstrate the non-extendibility of α to the region $|\alpha| \geq 1$, and illustrate the relationship to the likelihood ratio criterion (1). In Section 4, we recast the non-extendibility of α in connection with the (non-)extendibility of a conditional ∇^e-geodesic. Finally, we give brief concluding remarks in Section 5.

2 Proof of Theorem 3

2.1 α-divergence for unnormalized measures

The α-divergence (3) is nonnegative when p and q are both probability measures. However, it may take negative values when either p or q are unnormalized. The natural extension of α-divergence to unnormalized measures on $\mathcal{A}$ is as follows [11, p. 71]: for $\alpha \neq \pm 1$

$$\tilde{D}^{(\alpha)}(p\|q) := \frac{4}{1-\alpha^2} \sum_{x \in \mathcal{A}} \left[\frac{1-\alpha}{2} p(x) + \frac{1+\alpha}{2} q(x) - p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}} \right], \tag{5}$$

and $\tilde{D}^{(\pm 1)}(p\|q) := \lim_{\alpha \to \pm 1} \tilde{D}^{(\alpha)}(p\|q)$, that is,

$$\tilde{D}^{(-1)}(p\|q) := \tilde{D}^{(1)}(q\|p) := \sum_{x \in \mathcal{A}} \left[q(x) - p(x) + p(x) \log \frac{p(x)}{q(x)} \right]. \tag{6}$$

The α-divergence $\tilde{D}^{(\alpha)}(p\|q)$ is nonnegative, and equals zero if and only if $p = q$. To see this, it suffices to rewrite it in the form

$$\tilde{D}^{(\alpha)}(p\|q) = \sum_{x \in \mathcal{A}} p(x) \tilde{f}^{(\alpha)}\left(\frac{q(x)}{p(x)}\right),$$

where

$$\tilde{f}^{(\alpha)}(t) := \begin{cases} \dfrac{4}{1-\alpha^2} \left[\dfrac{1-\alpha}{2} + \dfrac{1+\alpha}{2} t - t^{\frac{1+\alpha}{2}} \right] & (\alpha \neq \pm 1) \\ 1 - t + t \log t & (\alpha = +1) \\ -1 + t - \log t & (\alpha = -1) \end{cases}$$

are strictly convex, nonnegative functions in t, with $\tilde{f}^{(\alpha)}(t) = 0$ if and only if $t = 1$.

Note that the functions $\tilde{f}^{(\alpha)}(t)$ satisfy the symmetry:

$$t\,\tilde{f}^{(\alpha)}(\frac{1}{t}) = \tilde{f}^{(-\alpha)}(t). \tag{7}$$

Further, they enjoy the following monotonicity in α:

$$\begin{aligned} 0 < t < 1 &\implies \tilde{f}^{(\alpha)}(t) \text{ is monotone decreasing in } \alpha \\ t > 1 &\implies \tilde{f}^{(\alpha)}(t) \text{ is monotone increasing in } \alpha. \end{aligned} \tag{8}$$

In fact, letting $\gamma := (-1+\alpha)/2$, we see that,

$$\frac{\partial}{\partial\alpha}\left[\frac{\partial\tilde{f}^{(\alpha)}}{\partial t}\right] = \frac{1}{2\gamma^2}\left[1 - t^{\gamma} + t^{\gamma}\log t^{\gamma}\right] \geq 0,$$

with equality if and only if $t = 1$. Thus $\partial\tilde{f}^{(\alpha)}/\partial t$ is monotone increasing in α. Let $\alpha_1 < \alpha_2$. Then for $0 < t < 1$

$$\tilde{f}^{(\alpha_1)}(t) = -\int_t^1 \frac{\partial\tilde{f}^{(\alpha_1)}}{\partial t}dt > -\int_t^1 \frac{\partial\tilde{f}^{(\alpha_2)}}{\partial t}dt = \tilde{f}^{(\alpha_2)}(t).$$

Similarly we can prove that $\tilde{f}^{(\alpha_1)}(t) < \tilde{f}^{(\alpha_2)}(t)$ for $t > 1$.

2.2 Basic lemma

For $x \in \mathcal{A}^*$, let $Km(x)$ be the monotone Kolmogorov complexity of x [2, p. 282]. It is known [2, p. 282, Theorem 4.5.4] that, given a computable probability measure P on $\mathcal{A}^\infty$, there is a constant C such that for all $\omega \in \mathcal{A}^\infty$ and $n \in \mathbb{N}$

$$Km(\omega^n) \leq -\log P(\omega^n) + C. \tag{9}$$

It is also known [2, p. 295, Corollary 4.5.3] that $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$ if and only if there is a constant $\tilde{C}$ such that for all n

$$Km(\omega^n) \geq -\log P(\omega^n) - \tilde{C}. \tag{10}$$

Let us introduce a randomness deficiency of $a \in \mathcal{A}^*$ relative to a computable probability measure P by

$$d(a|P) := -\log P(a) - Km(a).$$

Then, $d(\omega^n|P)$ is bounded from below for all $\omega \in \mathcal{A}^\infty$, and is bounded from above if and only if $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$.

A function $P : \mathcal{A}^* \to \mathbb{R}^+$ is called a *semimeasure* if it satisfies $P(\lambda) \leq 1$ and $P(a) \geq \sum_{x\in\mathcal{A}} P(ax)$ for all $a \in \mathcal{A}^*$. The randomness deficiency $d(a|P)$ can be extended formally to a computable semimeasure P, and we *define* $\omega \in \mathcal{A}^\infty$ to be random relative to a computable semimeasure P if $\sup_n d(\omega^n|P) < \infty$. Note that (9) holds also for a computable semimeasure P. To see this, let $C(\omega^{i-1}) := \sum_{\omega_i\in\mathcal{A}} P(\omega_i|\omega^{i-1})$ $(= \sum_{\omega_i\in\mathcal{A}} P(\omega^{i-1}\omega_i)/P(\omega^{i-1}) \leq 1)$ and let $\hat{P}(\omega_i|\omega^{i-1}) := P(\omega_i|\omega^{i-1})/C(\omega^{i-1})$. Then $\hat{P}(\omega^n) := \prod_{i=1}^n \hat{P}(\omega_i|\omega^{i-1}) = P(\omega^n)/\prod_{i=1}^n C(\omega^{i-1})$ is a computable probability measure on $\mathcal{A}^n$, and

$$-\log P(\omega^n) = -\log\hat{P}(\omega^n) - \log\left(\prod_{i=1}^n C(\omega^{i-1})\right) \geq -\log\hat{P}(\omega^n) \geq Km(\omega^n) - C.$$

Lemma 4. *Suppose P and Q are computable semimeasures. Then for all computable $\alpha \in \mathbb{R}$, there is a constant C such that for all $n \in \mathbb{N}$*

$$\frac{1-\alpha}{2}\, d(\omega^n|P) + \frac{1+\alpha}{2}\, d(\omega^n|Q) \geq \frac{1-\alpha^2}{4} \sum_{i=1}^{n} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) - C. \tag{11}$$

We regard (11) to be true if the left-hand side takes the indefinite form $\infty - \infty$.

Proof. When $\alpha = \pm 1$, (11) is reduced to (9) (under the convention that $0 \times \infty = 0$). Thus we need only treat the case when $\alpha \neq \pm 1$.

We first assume that $P(\omega^n) > 0$ and $Q(\omega^n) > 0$ for all n. Letting

$$R(\,\cdot\,|\omega^{i-1}) := \frac{1}{Z(\omega^{i-1})}\, P(\,\cdot\,|\omega^{i-1})^{\frac{1-\alpha}{2}} Q(\,\cdot\,|\omega^{i-1})^{\frac{1+\alpha}{2}},$$

with

$$Z(\omega^{i-1}) := \sum_{y \in \mathcal{A}} P(y|\omega^{i-1})^{\frac{1-\alpha}{2}} Q(y|\omega^{i-1})^{\frac{1+\alpha}{2}}$$

the normalization, we introduce a one-parameter family of probability measures on $\mathcal{A}^n$ as follows:

$$R(\omega^n) := \prod_{i=1}^{n} R(\omega_i|\omega^{i-1}) = \frac{1}{\prod_{i=1}^{n} Z(\omega^{i-1})}\, P(\omega^n)^{\frac{1-\alpha}{2}} Q(\omega^n)^{\frac{1+\alpha}{2}} = \frac{P(\omega^n)}{\prod_{i=1}^{n} Z(\omega^{i-1})} \left(\frac{Q(\omega^n)}{P(\omega^n)}\right)^{\frac{1+\alpha}{2}}. \tag{12}$$

Thus

$$-\log R(\omega^n) = -\log P(\omega^n) - \frac{1+\alpha}{2}\,(d(\omega^n|P) - d(\omega^n|Q)) + \sum_{i=1}^{n} \log Z(\omega^{i-1}).$$

According to (9), on the other hand, there is a constant C such that for all n,

$$-\log R(\omega^n) \geq Km(\omega^n) - C = -d(\omega^n|P) - \log P(\omega^n) - C.$$

By combining these two relations, we have

$$\frac{1-\alpha}{2}\, d(\omega^n|P) + \frac{1+\alpha}{2}\, d(\omega^n|Q) \geq -\sum_{i=1}^{n} \log Z(\omega^{i-1}) - C.$$

Since

$$D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) = \frac{4}{1-\alpha^2}(1 - Z(\omega^{i-1})),$$

we have

$$\begin{aligned} \log Z(\omega^{i-1}) \;&=\; \log\left[1 - \frac{1-\alpha^2}{4} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big)\right] \\ &\leq\; -\frac{1-\alpha^2}{4} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big). \end{aligned} \tag{13}$$

Here we used the inequality $\log(1+t) \leq t$. The assertion now follows immediately.

We next assume that $P(\omega^n) = 0$ or $Q(\omega^n) = 0$ for some n. In this case, (11) is reduced to either an inequality between infinities which is naturally regarded to be true, or an inequality having the indefinite form $\infty - \infty$ which is also regarded to be true because of the convention stated just after the assertion. The proof of Lemma 4 is completed. □

Corollary 5. *Suppose P and Q are computable semimeasures.*
(i) *For $\alpha > -1$, there is a constant C such that for all n*

$$d(\omega^n|Q) \geq \frac{1-\alpha}{2}\sum_{i=1}^{n} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) - \frac{1-\alpha}{1+\alpha}\,d(\omega^n|P) - \frac{2C}{1+\alpha}.$$

(ii) *For $\alpha < -1$, there is a constant C such that for all n*

$$d(\omega^n|Q) \leq \frac{1-\alpha}{2}\sum_{i=1}^{n} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) - \frac{1-\alpha}{1+\alpha}\,d(\omega^n|P) - \frac{2C}{1+\alpha}.$$

Remark 6. *Let us observe the relation:*

$$\tilde{D}^{(\alpha)}(p\|q) - D^{(\alpha)}(p\|q) = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}\left(\sum_{x\in\mathcal{A}} p(x) - 1\right) + \frac{1+\alpha}{2}\left(\sum_{x\in\mathcal{A}} q(x) - 1\right)\right],$$

where $\sum_{x\in\mathcal{A}} p(x) \leq 1$ and $\sum_{x\in\mathcal{A}} q(x) \leq 1$. If $-1 < \alpha < 1$, then $\tilde{D}^{(\alpha)}(p\|q) - D^{(\alpha)}(p\|q) \leq 0$. Thus Corollary 5 (i) can be modified as

$$d(\omega^n|Q) \geq \frac{1-\alpha}{2}\sum_{i=1}^{n} \tilde{D}^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) - \frac{1-\alpha}{1+\alpha}\,d(\omega^n|P) - \frac{2C}{1+\alpha}.$$

If $\alpha < -1$, on the other hand, and q is a probability measure, then $\tilde{D}^{(\alpha)}(p\|q) - D^{(\alpha)}(p\|q) \geq 0$. Therefore, if $Q(\,\cdot\,|\omega^{i-1})$ are probability measures on $\mathcal{A}$, Corollary 5 (ii) can be modified as

$$d(\omega^n|Q) \leq \frac{1-\alpha}{2}\sum_{i=1}^{n} \tilde{D}^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) - \frac{1-\alpha}{1+\alpha}\,d(\omega^n|P) - \frac{2C}{1+\alpha}.$$

2.3 Proof of 'only if' part of Theorem 3

That $Q(\omega^n) \neq 0$ for all n follows immediately from (10). On the other hand, since $\omega \in \mathrm{R}_{\mathrm{ML}}(P) \cap \mathrm{R}_{\mathrm{ML}}(Q)$, both $d(\omega^n|P)$ and $d(\omega^n|Q)$ are bounded from above. It then follows from Lemma 4 that

$$\sum_{i=1}^{\infty} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) < \infty$$

for all computable $\alpha \in (-1, 1)$. Moreover, since computable α's are dense in the interval $(-1, 1)$, and the α-divergence is monotone in α (see Appendix B), the above convergence holds for any real number $\alpha \in (-1, 1)$.

2.4 Proof of 'if' part of Theorem 3

We need to show that $\sup_n d(\omega^n|Q) < \infty$. In view of Corollary 5 (ii) or Remark 6, one might expect that the assumption

$$\sum_{i=1}^{\infty} \tilde{D}^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) < \infty, \qquad -1 < \exists\alpha < 1$$

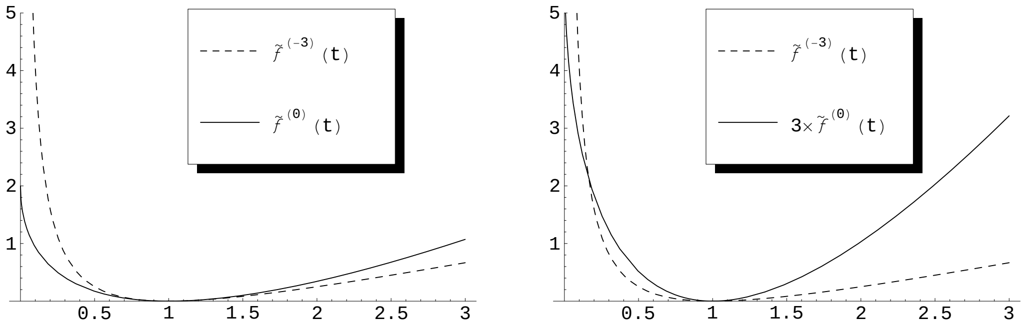

Figure 1: Whatever constant $K > 1$ we may choose, we cannot have $\tilde{f}^{(-3)}(t) \le K\tilde{f}^{(0)}(t)$ for $\forall t > 0$, because $\lim_{t\downarrow 0} \tilde{f}^{(0)}(t) = 2$ and $\lim_{t\downarrow 0} \tilde{f}^{(-3)}(t) = +\infty$. However, given a positive constant ε, it is possible to have $\tilde{f}^{(-3)}(t) \le K\tilde{f}^{(0)}(t)$ for $\forall t \ge \varepsilon$. This figure demonstrates the case when $K = 3$.

could imply that

$$\sum_{i=1}^{\infty} \tilde{D}^{(\beta)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) < \infty, \qquad \exists\beta < -1.$$

However, this is not true, as will be exemplified in Section 3.1. In fact, since

$$\lim_{t\downarrow 0} \tilde{f}^{(\alpha)}(t) = \frac{2}{1+\alpha}, \qquad \lim_{t\downarrow 0} \tilde{f}^{(\beta)}(t) = +\infty$$

for $\beta < -1 < \alpha$, there is no constant $K > 1$ that satisfies $\tilde{f}^{(\beta)}(t) \le K\tilde{f}^{(\alpha)}(t)$ for all $t > 0$. To surmount this difficulty, we need a device to restrict ourselves to a certain region $t \ge \varepsilon$ (> 0) in which $\tilde{f}^{(\beta)}(t) \le K\tilde{f}^{(\alpha)}(t)$ actually holds, see Figure 1.

We realize this program (with $\varepsilon = 1/3$) by introducing a reference probability measure $\overline{Q}$ and a semimeasure $\overline{P}$ as follows [7]:

(i) Let $\{A(a) \,|\, a \in \mathcal{A}^*\}$ be a computable family of subsets of $\mathcal{A}$ such that

$$Q(x|a) \le \frac{1}{2}\, P(x|a) \tag{14}$$

for $x \in A(a)$, and

$$Q(x|a) \ge \frac{1}{3}\, P(x|a) \tag{15}$$

for $x \notin A(a)$. Note that (15) cannot be replaced by the negation of (14) in general because the relation (14) is not always computable. Also it should be remarked that for a computable P, the set $\{a \in \mathcal{A}^* \,|\, P(a) = 0\}$ is tacitly assumed to be decidable [7].

(ii) For $a \in \mathcal{A}^*$ such that $Q(a) \ne 0$

$$\overline{Q}(x|a) := \begin{cases} \dfrac{P(x|a)}{2}, & x \in A(a) \\[2ex] Q(x|a)\cdot C(a), & x \notin A(a) \end{cases}$$

where $C(a) \in (0,1]$ is the normalization to assure that $\sum_x \overline{Q}(x|a) = 1$, that is,

$$C(a) := \frac{1 - \frac{P(A(a)|a)}{2}}{Q(A(a)^c|a)} \quad \left(\leq \frac{1 - Q(A(a)|a)}{Q(A(a)^c|a)} = 1 \right).$$

(iii) $\overline{P}(x|a) := P(x|a) \cdot C(a)$. This is actually a semimeasure since

$$\sum_{x \in \mathcal{A}} \overline{P}(ax) = \sum_{x \in \mathcal{A}} \overline{P}(x|a)\, \overline{P}(a) = C(a)\, \overline{P}(a) \leq \overline{P}(a).$$

Note that $\overline{P}$ and $\overline{Q}$ are both computable. According to the following chain decomposition of the likelihood ratio:

$$\frac{Q(\omega^n)}{P(\omega^n)} = \frac{\overline{P}(\omega^n)}{P(\omega^n)} \cdot \frac{\overline{Q}(\omega^n)}{\overline{P}(\omega^n)} \cdot \frac{Q(\omega^n)}{\overline{Q}(\omega^n)}$$

we subsequently prove that ω is random relative to $\overline{P}$ (Step 1), relative to $\overline{Q}$ (Step 2), and finally relative to Q (Step 3).

Step 0. Before proceeding to the proof, we make the following remark.

Lemma 7. *If $\sum_{i=1}^{\infty} \tilde{D}^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) < \infty$ for some $\alpha \in (-1,1)$, then*

$$\sum_{i=1}^{\infty} P(A(\omega^{i-1})|\omega^{i-1}) < \infty.$$

Proof. Recall that $\tilde{f}^{(\alpha)}(t)$ is nonnegative for all $t \in \mathbb{R}^+$, and is monotone decreasing in t for $0 < t < 1$ (cf. Figure 1). Since

$$x \in A(\omega^{i-1}) \implies \frac{Q(x|\omega^{i-1})}{P(x|\omega^{i-1})} \leq \frac{1}{2}$$

we have

$$\begin{aligned}
\tilde{D}^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) &= \sum_{x \in \mathcal{A}} P(x|\omega^{i-1}) \tilde{f}^{(\alpha)} \left(\frac{Q(x|\omega^{i-1})}{P(x|\omega^{i-1})} \right) \\
&\geq \sum_{x \in A(\omega^{i-1})} P(x|\omega^{i-1}) \tilde{f}^{(\alpha)} \left(\frac{Q(x|\omega^{i-1})}{P(x|\omega^{i-1})} \right) \\
&\geq \sum_{x \in A(\omega^{i-1})} P(x|\omega^{i-1}) \tilde{f}^{(\alpha)} \left(\frac{1}{2} \right) \\
&= \tilde{f}^{(\alpha)} \left(\frac{1}{2} \right) P(A(\omega^{i-1})|\omega^{i-1}).
\end{aligned}$$

Since $\tilde{f}^{(\alpha)}(1/2) > 0$, the assertion was verified. $\square$

Step 1. We need only show that

$$\lim_i \frac{\overline{P}(\omega^{i-1})}{P(\omega^{i-1})} = \prod_{i=1}^{\infty} C(\omega^{i-1}) \neq 0.$$

Observe that

$$C(\omega^{i-1}) \geq 1 - \frac{P(A(\omega^{i-1})|\omega^{i-1})}{2} \quad (> 0)$$

and that due to Lemma 7,

$$\sum_{i=1}^{\infty} \frac{P(A(\omega^{i-1})|\omega^{i-1})}{2} < \infty.$$

As a consequence

$$\prod_{i=1}^{\infty} C(\omega^{i-1}) \geq \prod_{i=1}^{\infty} \left(1 - \frac{P(A(\omega^{i-1})|\omega^{i-1})}{2}\right) > 0.$$

Here we have used the following well-known fact [10, p. 40]: for $0 \leq q_i < 1$,

$$\sum_{i=1}^{\infty} q_i < \infty \quad \Longleftrightarrow \quad \prod_{i=1}^{\infty} (1 - q_n) > 0.$$

Step 2. Fix a computable $\beta < -1$ arbitrarily. Then due to Remark 6, there is a constant C such that for all n

$$d(\omega^n|\overline{Q}) \leq \frac{1-\beta}{2} \sum_{i=1}^{n} \tilde{D}^{(\beta)}\big(\overline{P}(\,\cdot\,|\omega^{i-1})\|\overline{Q}(\,\cdot\,|\omega^{i-1})\big) - \frac{1-\beta}{1+\beta} d(\omega^n|\overline{P}) - \frac{2C}{1+\beta}.$$

We know from Step 1 that $\omega \in \mathrm{R_{ML}}(\overline{P})$, so that $\sup_n d(\omega^n|\overline{P}) < \infty$. It then suffices to prove that

$$\sum_{i=1}^{\infty} \tilde{D}^{(\beta)}\big(\overline{P}(\,\cdot\,|\omega^{i-1})\|\overline{Q}(\,\cdot\,|\omega^{i-1})\big) < \infty.$$

A crucial observation is that for $-1 < \alpha < 1$, there is a constant $K^{(\alpha,\beta)} > 1$ such that

$$t \geq \frac{1}{3} \implies \tilde{f}^{(\beta)}(t) \leq K^{(\alpha,\beta)} \tilde{f}^{(\alpha)}(t), \tag{16}$$

see Figure 1. In fact, let

$$K^{(\alpha,\beta)} := \sup\left\{ \frac{\tilde{f}^{(\beta)}(t)}{\tilde{f}^{(\alpha)}(t)} \;\middle|\; t \in [\frac{1}{3}, 1) \right\},$$

which is finite because $\lim_{t\to 1} \tilde{f}^{(\beta)}(t)/\tilde{f}^{(\alpha)}(t) = 1$. Then $\tilde{f}^{(\beta)}(t) \leq K^{(\alpha,\beta)} \tilde{f}^{(\alpha)}(t)$ for $1/3 \leq t < 1$. On the other hand, due to the monotonicity (8), we have $\tilde{f}^{(\beta)}(t) \leq \tilde{f}^{(\alpha)}(t) \leq K^{(\alpha,\beta)} \tilde{f}^{(\alpha)}(t)$ for $t \geq 1$, proving (16).

Further, let

$$L^{(\beta)} := \sup\left\{ \tilde{f}^{(-\beta)}(t) \;\middle|\; t \in (0, 2] \right\},$$

which is finite because $-\beta > 1$. Then

$$\begin{aligned}
&\tilde{D}^{(\beta)}\left(\overline{P}(\,\cdot\,|\omega^{i-1})\|\overline{Q}(\,\cdot\,|\omega^{i-1})\right) \\
&= \left(\sum_{x\in A(\omega^{i-1})} + \sum_{x\notin A(\omega^{i-1})}\right) \overline{P}(x|\omega^{i-1}) \tilde{f}^{(\beta)}\left(\frac{\overline{Q}(x|\omega^{i-1})}{\overline{P}(x|\omega^{i-1})}\right) \\
&= \sum_{x\in A(\omega^{i-1})} P(x|\omega^{i-1})C(\omega^{i-1})\tilde{f}^{(\beta)}\left(\frac{P(x|\omega^{i-1})/2}{P(x|\omega^{i-1})C(\omega^{i-1})}\right) \\
&\quad + \sum_{x\notin A(\omega^{i-1})} P(x|\omega^{i-1})C(\omega^{i-1})\tilde{f}^{(\beta)}\left(\frac{Q(x|\omega^{i-1})C(\omega^{i-1})}{P(x|\omega^{i-1})C(\omega^{i-1})}\right) \\
&= \frac{\tilde{f}^{(-\beta)}\left(2C(\omega^{i-1})\right)}{2} \sum_{x\in A(\omega^{i-1})} P(x|\omega^{i-1}) + C(\omega^{i-1}) \sum_{x\notin A(\omega^{i-1})} P(x|\omega^{i-1})\tilde{f}^{(\beta)}\left(\frac{Q(x|\omega^{i-1})}{P(x|\omega^{i-1})}\right) \\
&\le \frac{L^{(\beta)}}{2} P(A(\omega^{i-1})|\omega^{i-1}) + K^{(\alpha,\beta)} \sum_{x\notin A(\omega^{i-1})} P(x|\omega^{i-1})\tilde{f}^{(\alpha)}\left(\frac{Q(x|\omega^{i-1})}{P(x|\omega^{i-1})}\right) \\
&\le \frac{L^{(\beta)}}{2} P(A(\omega^{i-1})|\omega^{i-1}) + K^{(\alpha,\beta)}\tilde{D}^{(\alpha)}\left(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\right).
\end{aligned}$$

In the third equality we have used the symmetry (7). Due to Lemma 7 and the assumption, we have

$$\begin{aligned}
&\sum_{i=1}^{\infty} \tilde{D}^{(\beta)}\left(\overline{P}(\,\cdot\,|\omega^{i-1})\|\overline{Q}(\,\cdot\,|\omega^{i-1})\right) \\
&\qquad \le \frac{L^{(\beta)}}{2}\sum_{i=1}^{\infty} P(A(\omega^{i-1})|\omega^{i-1}) + K^{(\alpha,\beta)}\sum_{i=1}^{\infty}\tilde{D}^{(\alpha)}\left(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\right) < \infty.
\end{aligned}$$

Step 3. We prove that

$$\lim_{n\to\infty}\frac{Q(\omega^n)}{\overline{Q}(\omega^n)} = \prod_{i=1}^{\infty}\frac{Q(\omega_i|\omega^{i-1})}{\overline{Q}(\omega_i|\omega^{i-1})} \neq 0.$$

Since

$$\frac{Q(\omega_i|\omega^{i-1})}{\overline{Q}(\omega_i|\omega^{i-1})} = \begin{cases} \dfrac{Q(\omega_i|\omega^{i-1})}{P(\omega_i|\omega^{i-1})/2} \ (\le 1), & \text{if } \omega_i \in A(\omega^{i-1}) \\[2ex] \dfrac{1}{C(\omega^{i-1})} \ (\ge 1), & \text{otherwise} \end{cases}$$

it suffices to show that $\omega_i \in A(\omega^{i-1})$ only a finite number of times.

Since $\omega \in \mathrm{R}_{\mathrm{ML}}(P)\cap\mathrm{R}_{\mathrm{ML}}(\overline{Q})$, it holds from (1) that

$$\lim_{n\to\infty}\frac{\overline{Q}(\omega^n)}{P(\omega^n)} = \prod_{i=1}^{\infty}\frac{\overline{Q}(\omega_i|\omega^{i-1})}{P(\omega_i|\omega^{i-1})} \quad \text{converges to a positive number.}$$

In particular

$$\lim_{n\to\infty} \frac{\overline{Q}(\omega_i|\omega^{i-1})}{P(\omega_i|\omega^{i-1})} = 1.$$

Now suppose that $\omega_i \in A(\omega^{i-1})$ infinitely often. Then $\overline{Q}(\omega_i|\omega^{i-1}) = P(\omega_i|\omega^{i-1})/2$ infinitely many times, so that

$$\liminf_{n\to\infty} \frac{\overline{Q}(\omega_i|\omega^{i-1})}{P(\omega_i|\omega^{i-1})} \le \frac{1}{2}.$$

This is a contradiction. $\square$

3 Examples

In order to get a better perspective on Theorem 3, we give some illustrative examples, putting emphasis on the relationship to the likelihood ratio criterion (1).

3.1 Non-extendibility to the boundary $\alpha = \pm 1$

Is it possible to extend Theorem 3 to $\alpha = \pm 1$ and beyond? The answer is negative. We show this by a counterexample.

Let P and Q be independent stochastic processes, $P = \prod P_n$ and $Q = \prod Q_n$, on $\{0,1\}^\infty$ defined by

$$P_n(0) := \frac{1}{2n^2}, \qquad Q_n(0) := e^{-n}.$$

Note that $P \overset{loc}{\sim} Q$. For these processes

$$\frac{1}{4} D^{(0)}(P_n\|Q_n) = 1 - \sum_{x\in\{0,1\}} \sqrt{P_n(x)Q_n(x)} \le 1 - \sum_{x\in\{0,1\}} P_n(x)Q_n(x) = \frac{1}{2n^2} + e^{-n} - \frac{e^{-n}}{n^2}.$$

Thus $\sum_{n=1}^\infty D^{(0)}(P_n\|Q_n) < \infty$, so that $P \sim Q$, and that $\mathrm{R_{ML}}(P) = \mathrm{R_{ML}}(Q)$ due to Theorem 2. On the other hand,

$$\begin{aligned} D^{(-1)}(P_n\|Q_n) = \tilde{D}^{(-1)}(P_n\|Q_n) \quad &\ge \quad Q_n(0) - P_n(0) + P_n(0)\log\frac{P_n(0)}{Q_n(0)} \\ &= \quad e^{-n} - \frac{1}{2n^2} + \frac{1}{2n^2}\log\left(\frac{e^n}{2n^2}\right) = O\left(\frac{1}{n}\right). \end{aligned}$$

Thus $\sum_{n=1}^\infty D^{(-1)}(P_n\|Q_n) = \infty = \sum_{n=1}^\infty D^{(+1)}(Q_n\|P_n)$. Similarly we can easily prove that $\sum_{n=1}^\infty D^{(\alpha)}(P_n\|Q_n) = \infty = \sum_{n=1}^\infty D^{(-\alpha)}(Q_n\|P_n)$ for $\alpha < -1$. As a consequence, the 'only if' part of Theorem 3 does not hold for $|\alpha| \ge 1$. In other words, Theorem 3 gives the best possible criterion in terms of α-divergence.

Let us recast this example in connection with the likelihood ratio test. Since $\mathrm{R_{ML}}(P) = \mathrm{R_{ML}}(Q)$, it follows from (1) that for all $\omega \in \mathrm{R_{ML}}(P)$

$$\lim_{n\to\infty} \frac{Q(\omega^n)}{P(\omega^n)} = \prod_{n=1}^\infty \frac{Q_n(\omega_n)}{P_n(\omega_n)} \quad \text{converges to a positive number.}$$

In particular

$$\lim_{n\to\infty} \frac{Q_n(\omega_n)}{P_n(\omega_n)} = 1.$$

How can this be consistent with the fact that

$$\lim_{n\to\infty} \frac{Q_n(0)}{P_n(0)} = 0?$$

The answer is that $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$ contains only a finite number of 0's. In fact, suppose ω contains infinitely many 0's. Then

$$\liminf_{n\to\infty} \frac{Q_n(\omega_n)}{P_n(\omega_n)} = 0.$$

This is a contradiction.

This observation illustrates that a randomness criterion cannot be too sensitive to the behavior $Q_n(x)/P_n(x) \to 0$ for some $x \in \mathcal{A}$. This is why our criterion is restricted to $-1 < \alpha < 1$, for which $\lim_{t\downarrow 0} \tilde{f}^{(\alpha)}(t) < \infty$.

3.2 On the condition $Q(\omega^n) \neq 0$ in Theorem 3

In Theorem 3, the condition that $Q(\omega^n) \neq 0$ $(\forall n)$ cannot be dispensed with. In fact, $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$ and $\sum_{n=1}^{\infty} D^{(0)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) < \infty$ together do not imply $Q(\omega^n) \neq 0$ $(\forall n)$.

Let P and Q be independent processes on $\{0,1\}^\infty$ defined by

$$P_n(0) = \frac{1}{2n^2}, \qquad Q_n(0) = 0.$$

This pair exhibits $Q \overset{loc}{\ll} P$, and

$$\frac{1}{4}\sum_{n=1}^{\infty} D^{(0)}(P_n \| Q_n) \le \sum_{n=1}^{\infty} \left[1 - \sum_{x\in\{0,1\}} P_n(x) Q_n(x)\right] = \sum_{n=1}^{\infty} \frac{1}{2n^2} < \infty.$$

Thus $Q \ll P$, and $\mathrm{R}_{\mathrm{ML}}(Q) \subset \mathrm{R}_{\mathrm{ML}}(P)$ due to Theorem 2. In fact, $\mathrm{R}_{\mathrm{ML}}(Q) = \{1^\infty\}$, and

$$\lim_{n\to\infty} \frac{Q(1^n)}{P(1^n)} = \frac{1}{\prod_{n=1}^{\infty}\left(1 - \frac{1}{2n^2}\right)},$$

which converges to a positive number. All the other elements ω $(\neq 1^\infty)$ in $\mathrm{R}_{\mathrm{ML}}(P)$ contain at least one 0's, and $Q(\omega^n) = 0$ eventually.

Note that the condition $Q(\omega^n) \neq 0$ $(\forall n)$ assures the existence of the reference measure $\overline{Q}(\omega^n)$ in Section 2.4.

3.3 Bienvenu-Merkle process

We next investigate somewhat peculiar pair of computable probability measures P and Q on $\{0,1\}^\infty$ which exhibit $P \sim Q$ and $\mathrm{R_{ML}}(P) \neq \mathrm{R_{ML}}(Q)$. Let

$$P(\omega^n) := \frac{1}{2^n}, \qquad Q(\omega^n) := r(\omega^n)P(\omega^n),$$

where the likelihood ratio $r(\omega^n)$ is defined by (a slight modification of) the Bienvenu and Merkle construction [5] as follows.

Let Ω be Chaitin's halting probability [2, p. 217], which may be identified with the binary infinite sequence after the decimal point. Note that $\Omega \in \mathrm{R_{ML}}(P)$. Since Ω is Δ^0_2-definable, we can choose a computable sequence $\{w_s\}_{s\geq 0}$ of words which converges bitwise to Ω. Here we assume that $\lim_s |w_s| = \infty$, that $|w_s| \leq 3^s$ for all s, and that $w_s \sqsubset \Omega$ (i.e., w_s is a prefix of Ω) for infinitely many s. For example, let $\{v_s\}_{s\geq 0}$ be a standard computable increasing sequence of positive rationals of the form $\sum_{i=0}^{s} 2^{-\ell_i}$ ($\ell_i \in \mathbb{N}$) which converges to the real number Ω. Identify each v_s with the binary sequence after the decimal point, terminating in '1'. We then take w_s to be the longest common prefix of v_s and v_{s+1}, unless it is longer than 3^s.

Step 0: Set $r(0) = r(1) = 1$.

Step s (≥ 1): We first define $r(u)$ for $|u| = 3^s$ by using w_{s-1}; we then define $r(u)$ for $3^{s-1} < |u| < 3^s$ by induction.

(i) for each $u \in \{0,1\}^{3^s}$ let $\overline{u}$ be the prefix of u of length 3^{s-1}, that is, $u = \overline{u}\,a$ with $\exists a \in \{0,1\}^{3^s - 3^{s-1}}$, and let

$$r(u) := \begin{cases} r(\overline{u}), & \text{if } w_{s-1} \not\sqsubset \overline{u} \\[2ex] \dfrac{1}{2}\, r(\overline{u}), & \text{if } w_{s-1} \sqsubset \overline{u} \text{ and } u \neq \overline{u}\,111\cdots 1 \\[2ex] x(\overline{u}), & \text{if } w_{s-1} \sqsubset \overline{u} \text{ and } u = \overline{u}\,111\cdots 1 \end{cases} \tag{17}$$

where $x(\overline{u})$ is determined in such a way that the average value of $\left\{ r(\overline{u}\,a) \;\middle|\; a \in \{0,1\}^{3^s - 3^{s-1}} \right\}$ for each $\overline{u}$ ($\sqsupset w_{s-1}$) of length 3^{s-1} is equal to $r(\overline{u})$. More precisely

$$\frac{(2^{3^s - 3^{s-1}} - 1)\, \dfrac{r(\overline{u})}{2} + x(\overline{u})}{2^{3^s - 3^{s-1}}} = r(\overline{u}),$$

so that

$$x(\overline{u}) = \left(\frac{2^{3^s - 3^{s-1}} + 1}{2} \right) r(\overline{u}).$$

(ii) for $3^{s-1} < |u| < 3^s$, set inductively (in a reverse direction) as

$$r(u) := \frac{r(u0) + r(u1)}{2}.$$

It should be noted that the recursion formula in (ii) is identical to the martingale condition:

$$r(u)P(u) = r(u0)P(u0) + r(u1)P(u1).$$

Therefore, due to the definition of $x(\overline{u})$ in (17), the induction (ii) can be extended consistently to $|u| = 3^{s-1}$.

Since $w_s \to \Omega$, the sequence $r(\omega^n)$ enjoys the following properties:

(a) $r(\Omega^n) \to 0$,

(b) for $\forall \omega \neq \Omega$, $r(\omega^n)$ will be a positive constant eventually.

As a consequence, for $Q(\omega^n) := r(\omega^n)P(\omega^n)$,

$$\left\{ \omega \;\middle|\; \inf_n \frac{Q(\omega^n)}{P(\omega^n)} = 0 \right\} = \{\Omega\}, \qquad \left\{ \omega \;\middle|\; \sup_n \frac{Q(\omega^n)}{P(\omega^n)} = \infty \right\} = \emptyset.$$

Therefore, due to the following characterization [9, p. 527, Theorem 2]: for $Q \overset{loc}{\ll} P$,

$$Q \ll P \iff Q\left(\limsup_n \frac{Q(\omega^n)}{P(\omega^n)} = \infty \right) = 0,$$

we have $P \sim Q$. Moreover, according to the randomness criterion (1) based on the likelihood ratio test, we have

$$\mathrm{R}_{\mathrm{ML}}(Q) = \mathrm{R}_{\mathrm{ML}}(P) \backslash \{\Omega\}.$$

Let us recast this result in the light of Theorem 3. We conceive of each element $\omega \in \{0,1\}^\infty$ as a path in the infinite binary tree. According to the assumption, there are infinitely many nodes w_s on the path Ω. Among those s (satisfying $w_s \sqsubset \Omega$), the node Ω^{3^s} falls into the third case in (17) at most finitely many times of s, since otherwise it would contradict the incompressibility of Ω. All the other s (satisfying $w_s \sqsubset \Omega$) fall into the second case in (17), and there is an i, $(3^{s-1} < i \le 3^s)$, for which

$$r(\Omega^i) \le \frac{1}{2} r(\Omega^{i-1}).$$

It follows that

$$\begin{aligned}
\tilde{D}^{(\alpha)}\big(P(\,\cdot\,|\Omega^{i-1}) \| Q(\,\cdot\,|\Omega^{i-1})\big) &= \sum_{x \in \{0,1\}} P(x|\Omega^{i-1}) \tilde{f}^{(\alpha)} \left(\frac{Q(x|\Omega^{i-1})}{P(x|\Omega^{i-1})} \right) \\
&\ge P(\Omega_i|\Omega^{i-1}) \tilde{f}^{(\alpha)} \left(\frac{Q(\Omega_i|\Omega^{i-1})}{P(\Omega_i|\Omega^{i-1})} \right) \\
&= \frac{1}{2} \tilde{f}^{(\alpha)} \left(\frac{r(\Omega^i)}{r(\Omega^{i-1})} \right) \\
&\ge \frac{1}{2} \tilde{f}^{(\alpha)} \left(\frac{1}{2} \right)
\end{aligned}$$

for infinitely many i's. We thus conclude that $\sum_i \tilde{D}^{(\alpha)}\big(P(\,\cdot\,|\Omega^{i-1}) \| Q(\,\cdot\,|\Omega^{i-1})\big) = \infty$, proving $\Omega \notin \mathrm{R}_{\mathrm{ML}}(Q)$.

For $\omega \neq \Omega$, on the other hand, it follows from (b) that

$$\frac{Q(x|\omega^{i-1})}{P(x|\omega^{i-1})} = \frac{r(\omega^{i-1}\,x)}{r(\omega^{i-1})} = 1, \qquad \forall x \in \{0,1\},$$

for all but finitely many i's. As a consequence, $\sum_i \tilde{D}^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) < \infty$ for $\forall\omega \in \mathrm{R}_{\mathrm{ML}}(P)\backslash\{\Omega\}$ and $\forall\omega \in \mathrm{R}_{\mathrm{ML}}(Q)$, proving that $\mathrm{R}_{\mathrm{ML}}(Q) = \mathrm{R}_{\mathrm{ML}}(P)\backslash\{\Omega\}$.

4 Extendibility of conditional ∇^e-geodesic

In this section, we recast the non-extendibility of the randomness criterion to $|\alpha| \geq 1$ from a different angle. Let us recall the probability measure $R(\omega^n)$ on $\mathcal{A}^n$ defined by (12), i.e.,

$$R(\omega^n) = \prod_{i=1}^{n} R(\omega_i|\omega^{i-1}) = \frac{P(\omega^n)}{\prod_{i=1}^{n} Z(\omega^{i-1})}\left(\frac{Q(\omega^n)}{P(\omega^n)}\right)^{\frac{1+\alpha}{2}},$$

with

$$Z(\omega^{i-1}) = \sum_{x\in\mathcal{A}} P(x|\omega^{i-1})^{\frac{1-\alpha}{2}} Q(x|\omega^{i-1})^{\frac{1+\alpha}{2}}.$$

Since the sequence $\{R(\omega^n)\}_n$ of probability measures enjoys the consistency:

$$\sum_{\omega_{n+1}\in\mathcal{A}} R(\omega^{n+1}) = \sum_{\omega_{n+1}\in\mathcal{A}} R(\omega_{n+1}|\omega^n)\prod_{i=1}^{n} R(\omega_i|\omega^{i-1}) = R(\omega^n),$$

it can be uniquely extended to $\mathcal{A}^\infty$, for which we use the same symbol R. When α needs to be specified, we denote it as $R^{(\alpha)}$, and correspondingly $Z^{(\alpha)}(\omega^{i-1})$ for $Z(\omega^{i-1})$.

Unless P and Q are both product measures, $R^{(\alpha)}(\omega^n)$ is *not* the genuine ∇^e-geodesic [11] connecting two measures $P(\omega^n)$ and $Q(\omega^n)$ on $\mathcal{A}^n$, since the normalization $\prod_{i=1}^{n} Z^{(\alpha)}(\omega^{i-1})$ is an $\mathcal{F}_{n-1}$-measurable random variable depending on ω^{n-1}. We may call $R^{(\alpha)}$ a *conditional* ∇^e*-geodesic* connecting P and Q. In this section, we investigate if $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$ is still random relative to $R^{(\alpha)}$.

4.1 Conservation of randomness for $-1 \leq \alpha \leq 1$

We first note that randomness is conserved under the ∇^e-convex combination of probability measures.

Proposition 8. *Let P and Q be computable probability measures, and let $\omega \in \mathrm{R}_{\mathrm{ML}}(P) \cap \mathrm{R}_{\mathrm{ML}}(Q)$. Then $\omega \in \mathrm{R}_{\mathrm{ML}}(R^{(\alpha)})$ for all computable $\alpha \in [-1,1]$.*

Proof. Recall that

$$-\log P(\omega^n) \leq Km(\omega^n) + \tilde{C}_P \quad \text{and} \quad -\log Q(\omega^n) \leq Km(\omega^n) + \tilde{C}_Q$$

The assertion is trivial for $\alpha = \pm 1$. For $\alpha \in (-1,1)$, it follows from Jensen's inequality that $Z(\omega^{i-1}) \leq 1$, so that

$$\begin{aligned} -\log R^{(\alpha)}(\omega^n) &= \frac{1-\alpha}{2}(-\log P(\omega^n)) + \frac{1+\alpha}{2}(-\log Q(\omega^n)) + \log\left(\prod_{i=1}^{n} Z^{(\alpha)}(\omega^{i-1})\right) \qquad (18)\\ &\leq Km(\omega^n) + \frac{1-\alpha}{2}\tilde{C}_P + \frac{1+\alpha}{2}\tilde{C}_Q. \end{aligned}$$

$\square$

Note that (18) forces to conclude that $\prod_{i=1}^{\infty} Z^{(\alpha)}(\omega^{i-1}) > 0$, since otherwise it contradicts (9). For $\alpha \in (-1,1)$, this is equivalent to $\sum_{i=1}^{\infty} D^{(\alpha)}\big(P(\,\cdot\,\|\omega^{i-1})\|Q(\,\cdot\,\|\omega^{i-1})\big) < \infty$, because $Z^{(\alpha)}(\omega^{i-1}) = 1 - \big((1-\alpha^2)/4\big)\, D^{(\alpha)}\big(P(\,\cdot\,\|\omega^{i-1})\|Q(\,\cdot\,\|\omega^{i-1})\big)$. This gives an alternative proof of the 'only if' part of Theorem 3.

4.2 Extendibility to $\alpha < -1$

The following theorem concerns the possibility of extending the conditional ∇^e-geodesic segment connecting P and Q beyond P, i.e. $\alpha < -1$, keeping $\omega \in \mathrm{R_{ML}}(P)$ still random relative to $R^{(\alpha)}$.

Theorem 9. *Let P and Q be computable probability measures, and let $\omega \in \mathrm{R_{ML}}(P)$. Then for computable $\alpha < -1$,*

$$\prod_{i=1}^{\infty} Z^{(\alpha)}(\omega^{i-1}) < \infty \implies \omega \in \mathrm{R_{ML}}(R^{(\alpha)}). \tag{19}$$

Proof. It follows from Jensen's inequality that $Z^{(\alpha)}(\omega^{i-1}) \geq 1$ for $\alpha < -1$. Further

$$\begin{aligned} -\log R^{(\alpha)}(\omega^n) &= \frac{1-\alpha}{2}(-\log P(\omega^n)) + \frac{1+\alpha}{2}(-\log Q(\omega^n)) + \log\left(\prod_{i=1}^{n} Z^{(\alpha)}(\omega^{i-1})\right) \\ &\leq Km(\omega^n) + \frac{1-\alpha}{2}\tilde{C}_P + \frac{1+\alpha}{2}(-C_Q) + \log\left(\prod_{i=1}^{n} Z^{(\alpha)}(\omega^{i-1})\right) \end{aligned}$$

Here we have used (9). Thus $\displaystyle\prod_{i=1}^{\infty} Z^{(\alpha)}(\omega^{i-1}) < \infty$ implies $\omega \in \mathrm{R_{ML}}(R^{(\alpha)})$. $\square$

The condition $\prod_{i=1}^{\infty} Z^{(\alpha)}(\omega^{i-1}) < \infty$ in Theorem 9 is closely related to Corollary 5 (ii). In fact, we have the following chain of implications: suppose $\omega \in \mathrm{R_{ML}}(P)$ and $\alpha < -1$, then

$$\sum_{i=1}^{\infty} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) < \infty \overset{(13)}{\implies} \prod_{i=1}^{\infty} Z^{(\alpha)}(\omega^{i-1}) < \infty \overset{(19)}{\implies} \omega \in \mathrm{R_{ML}}(R^{(\alpha)}). \tag{20}$$

On the other hand, we see from the identity

$$\prod_{i=1}^{n} Z^{(\alpha)}(\omega^{i-1}) = \left(\frac{R^{(\alpha)}(\omega^n)}{P(\omega^n)}\right)^{-1} \left(\frac{Q(\omega^n)}{P(\omega^n)}\right)^{\frac{1+\alpha}{2}}$$

that

$$\prod_{i=1}^{\infty} Z^{(\alpha)}(\omega^{i-1}) < \infty \text{ and } \omega \in \mathrm{R_{ML}}(P) \cap \mathrm{R_{ML}}(R^{(\alpha)}) \implies \omega \in \mathrm{R_{ML}}(P) \cap \mathrm{R_{ML}}(Q).$$

Putting these together, for $\omega \in \mathrm{R_{ML}}(P)$ and $\alpha < -1$,

$$\sum_{i=1}^{\infty} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) < \infty \implies \omega \in \mathrm{R_{ML}}(Q).$$

This is essentially equivalent to Corollary 5 (ii).

Although the convergence of $\sum_{i=1}^{\infty} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big)$ for $\alpha < -1$ is too strong a requirement for a randomness criterion (Section 3.1), (20) suggests that the convergence of $\prod_{i=1}^{\infty} Z^{(\alpha)}(\omega^{i-1})$ for $\alpha < -1$ might be worth investigation. Since

$$\frac{\partial^2}{\partial \alpha^2} \log Z(\omega^{i-1}) = \frac{1}{Z(\omega^{i-1})} \sum_{y \in \mathcal{A}} P(y|\omega^{i-1}) \left(\frac{Q(y|\omega^{i-1})}{P(y|\omega^{i-1})} \right)^{\frac{1+\alpha}{2}} \left(\frac{1}{2} \log \frac{Q(y|\omega^{i-1})}{P(y|\omega^{i-1})} - \frac{Z(\omega^{i-1})'}{Z(\omega^{i-1})} \right)^2 > 0,$$

the functions $\alpha \mapsto \log Z^{(\alpha)}(\omega^{i-1})$ are convex for all i, and so is

$$\psi(\alpha) := \sum_{i=1}^{\infty} \log Z^{(\alpha)}(\omega^{i-1}) = \log \left(\prod_{i=1}^{\infty} Z^{(\alpha)}(\omega^{i-1}) \right).$$

Further, since $Z^{(-1)}(\omega^{i-1}) = 1$, it holds that $\psi(-1) = 0$ and

$$\left.\frac{\partial \psi}{\partial \alpha}\right|_{\alpha=-1} = \sum_{i=1}^{\infty} \left.\frac{\partial}{\partial \alpha} \log Z^{(\alpha)}(\omega^{i-1})\right|_{\alpha=-1} = -\frac{1}{2} \sum_{i=1}^{\infty} D^{(-1)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big).$$

It is thus crucial to investigate if $\sum_{i=1}^{\infty} D^{(-1)}\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) < \infty$.

Suppose this was true for a given pair P, Q of probability measures. Then one could extend $\psi(\alpha)$ continuously to $\alpha < -1$, to conclude that the 'if' part of Theorem 3 is proved directly from Theorem 9, and that the conditional ∇^e-geodesic is extended to outside the segment. Unfortunately, this is not always the case, as the counterexample in Section 3.1 shows.

5 Concluding remarks

Vovk's randomness criterion was extended to α-divergences for $\alpha \in (-1, 1)$. It was also shown that the criterion cannot be extended to $|\alpha| \geq 1$. In this sense, Theorem 3 characterizes all the possible randomness criteria in terms of α-divergence.

Criteria of randomness are closely connected with criteria of absolute continuity and singularity of probability measures. For example, as stated in [7], Kabanov-Liptser-Shiryaev's criterion [12] (see also [9, Chapter VII, Section 6.3, Theorem 4]) of absolute continuity and singularity of probability measures on $\mathcal{A}^{\infty}$ can be derived from Vovk's criterion of randomness. Similarly, our criteria of randomness naturally lead to α-divergence extension of Kabanov-Liptser-Shiryaev's criterion. A special case of this extension is discussed in Appendix A from a different point of view.

It should be noted that one could prove Theorem 3 by combining Theorem 1 with the following fact: If $\sum_n D^{(\alpha)}(p_n\|q_n) < \infty$ for some $\alpha \in (-1, 1)$, then $\sum_n D^{(\beta)}(p_n\|q_n) < \infty$ for all $\beta \in (-1, 1)$. (See Appendix B.) Nevertheless, such a way of understanding Theorem 3, in which a certain value of α plays a crucial role, is somewhat awkward. In contrast, the proof presented in Section 2 does not favor any specific value of $\alpha \in (-1, 1)$ and is mathematically natural. Moreover, it makes clear the distinction between the underlying mechanism of the cases $|\alpha| < 1$ and $|\alpha| \geq 1$.

Acknowledgments

I am particularly grateful to Phil Dawid for his hospitality during my visit to UCL. I also would like to thank Volodya Vovk for stimulating discussions, Alexander Shen for bringing paper [5] to my attention, and the anonymous referees for valuable comments.

Appendix

A A generalization of Kakutani's Theorem

In this appendix, we derive α-divergence versions of Kakutani's criterion from a different point of view.

Theorem 10. *Let $X_1, X_2, \ldots$ be independent nonnegative random variables, each of mean 1. Define $M_n := X_1 X_2 \cdots X_n$, with $M_0 := 1$. Then M is a nonnegative martingale, so that $M_\infty := \lim M_n$ exists a.s. Given $p > 1$, let $a_n := E[X_n^{1/p}]$. The following five statements are equivalent:*
(i) $E[M_\infty] = 1$
(ii) $M_n \to M_\infty$ *in* $\mathcal{L}^1$
(iii) M *is uniformly integrable*
(iv) $\prod_n a_n > 0$
(v) $\sum_n (1 - a_n) < \infty$
If one (then every one) of the above five statements fails to hold, then $P(M_\infty = 0) = 1$.

Proof. We need only show that (iv) implies that $(M_n)_n$ is $\mathcal{L}^1$-dominated [10, Section 14.12]. Let

$$N_n := \frac{X_1^{1/p}}{a_1} \cdots \frac{X_n^{1/p}}{a_n}.$$

Note that $0 < a_n \le 1$ by Jensen's inequality, and $(N_n)_n$ is a martingale. Suppose $\prod_n a_n > 0$. Then

$$E[N_n^p] = \frac{1}{(a_1 \cdots a_n)^p} \le \frac{1}{(\prod_k a_k)^p} < \infty,$$

so that $(N_n)_n$ is bounded in $\mathcal{L}^p$. Since

$$N_n^p = \frac{M_n}{(a_1 \cdots a_n)^p} \ge M_n,$$

it holds from Doob's $\mathcal{L}^p$ inequality [10, Section 14.11] that,

$$E[\sup_n |M_n|] \le E[\sup_n |N_n^p|] \le q^p \sup_n E[N_n^p] < \infty,$$

where $(1/p) + (1/q) = 1$. As a consequence, $(M_n)_n$ is dominated by $M^* := \sup_n |M_n| \in \mathcal{L}^1$. □

If we conceive of X_n in Theorem 10 as the likelihood ratio Q_n/P_n of two probability measures satisfying $Q_n \ll P_n$ for all n, then

$$1 - a_n = 1 - E_P\left[\left(\frac{Q_n}{P_n}\right)^{1/p}\right] = 1 - \sum_{x \in \mathcal{A}} P_n(x)^{1-1/p} Q_n(x)^{1/p},$$

and the condition (v) is equivalent to

$$\sum_n D^{(\alpha)}(P_n \| Q_n) < \infty, \tag{21}$$

where $\alpha := (2/p) - 1 \in (-1, 1)$. Since $Q \ll P$ if and only if $(M_n)_n$ is uniformly integrable [10, Section 14.16], it holds that $Q \ll P$ if and only if $\sum_n D^{(\alpha)}(P_n \| Q_n) < \infty$ for $\exists \alpha \in (-1, 1)$; otherwise $P \perp Q$. The original Kakutani theorem corresponds to the case when $p = 2$ (or $\alpha = 0$).

B Monotonicity of α-divergence in α

Lemma 11. *Let $\{p_n\}_n$ and $\{q_n\}_n$ be sequences of probability measures. If $\sum_n D^{(\alpha)}(p_n \| q_n) < \infty$ for some $\alpha \in (-1, 1)$, then $\sum_n D^{(\beta)}(p_n \| q_n) < \infty$ for all $\beta \in (-1, 1)$.*

Proof. Let us rewrite the α-divergence as

$$D^{(\alpha)}(p \| q) = \frac{2}{1 - \alpha} \sum_{x \in \mathcal{A}} p(x) f^{(\alpha)} \left(\frac{q(x)}{p(x)} \right),$$

where

$$f^{(\alpha)}(t) := \frac{2}{1 + \alpha}(1 - t^{\frac{1+\alpha}{2}}).$$

Since

$$\frac{\partial}{\partial \gamma} \left(\frac{1 - t^\gamma}{\gamma} \right) = \frac{1}{\gamma^2}(-t^\gamma \log t^\gamma - 1 + t^\gamma) \le 0,$$

we see that $f^{(\alpha)}(t)$ is monotone decreasing in α for each $t > 0$. As a consequence,

$$\sum_{x \in \mathcal{A}} p(x) f^{(\alpha)} \left(\frac{q(x)}{p(x)} \right) \ge \sum_{x \in \mathcal{A}} p(x) f^{(\beta)} \left(\frac{q(x)}{p(x)} \right)$$

for all $\beta \ge \alpha$. Now suppose that $\sum_n D^{(\alpha)}(p_n \| q_n) < \infty$ for some $\alpha \in (-1, 1)$. Since this is equivalent to

$$\sum_n \left\{ \sum_{x \in \mathcal{A}} p_n(x) f^{(\alpha)} \left(\frac{q_n(x)}{p_n(x)} \right) \right\} < \infty,$$

we conclude from the above monotonicity that

$$\sum_n \left\{ \sum_{x \in \mathcal{A}} p_n(x) f^{(\beta)} \left(\frac{q_n(x)}{p_n(x)} \right) \right\} < \infty$$

for all $\beta \ge \alpha$, and that $\sum_n D^{(\beta)}(p_n \| q_n) < \infty$ for all $\beta \in [\alpha, 1)$.

We next rewrite the α-divergence as

$$D^{(\alpha)}(p \| q) = \frac{2}{1 + \alpha} \sum_{x \in \mathcal{A}} q(x) f^{(-\alpha)} \left(\frac{p(x)}{q(x)} \right).$$

Since, for each $t > 0$, $f^{(-\alpha)}(t)$ is monotone increasing in α, it holds that

$$\sum_{x\in\mathcal{A}} q(x) f^{(-\beta)}\left(\frac{p(x)}{q(x)}\right) \le \sum_{x\in\mathcal{A}} q(x) f^{(-\alpha)}\left(\frac{p(x)}{q(x)}\right)$$

for all $\beta \le \alpha$. Since $\sum_n D^{(\alpha)}(p_n \| q_n) < \infty$ is equivalent to

$$\sum_n \left\{ \sum_{x\in\mathcal{A}} q_n(x) f^{(-\alpha)}\left(\frac{p_n(x)}{q_n(x)}\right) \right\} < \infty,$$

we conclude from the monotonicity that

$$\sum_n \left\{ \sum_{x\in\mathcal{A}} q_n(x) f^{(-\beta)}\left(\frac{p_n(x)}{q_n(x)}\right) \right\} < \infty$$

for all $\beta \le \alpha$, and that $\sum_n D^{(\beta)}(p_n \| q_n) < \infty$ for all $\beta \in (-1, \alpha]$. $\square$

References

[1] P. Martin-Löf, "The definition of random sequences," Information and Control **9**, pp. 602-619 (1966).

[2] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd ed. (Springer, NY, 1997).

[3] P. Martin-Löf, *Notes on Constructive Mathematics* (Almqvist & Wiksell, Stockholm, 1970).

[4] V. Vovk, Private communication.

[5] L. Bienvenu and W. Merkle, "Effective randomness for computable probability measures," Electronic Notes in Theoretical Computer Science **167**, pp. 117-130 (2007).

[6] H. Takahashi, "Algorithmic randomness and monotone complexity on product space," preprint.

[7] V. Vovk, "On a randomness criterion," Soviet Math. Dokl. **35**, pp. 656-660 (1987).

[8] S. Kakutani, "On equivalence of infinite product measures," Ann. Math. **49**, pp. 214-224 (1948).

[9] A. N. Shiryaev, *Probability*, 2nd ed. (Springer, NY, 1996).

[10] D. Williams, *Probability with Martingales* (Cambridge University Press, Cambridge, 1991).

[11] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monograph, Vol. 191 (AMS and Oxford University Press, 2000).

[12] Y. M. Kabanov, R. S. Liptser, and A. N. Shiryaev, "On the question of absolute continuity and singularity of probability measures," Mathematics of the USSR—Sbornik **33**, pp. 203-221 (1977).