

CAM 1208

The measure-theoretic aspects of entropy, Part I

P.R. Masani

Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260, United States

Received 9 November 1991

Abstract

Masani, P.R., The measure-theoretic aspects of entropy, Part I, Journal of Computational and Applied Mathematics 40 (1992) 215–232.

Let P be a countably additive probability measure and μ be any $[0, \infty]$ -valued countably additive measure, both on a σ -algebra $\mathfrak{A}$ over a space Ω . We define a (P -, μ -dependent) countably additive measure H on the δ -ring $\mathfrak{A}_\mu = \{A: A \in \mathfrak{A} \text{ and } \mu(A) < \infty\}$ with values in $(-\infty, +\infty]$ with the property that in case $\Omega \in \mathfrak{A}_\mu$, $H(\Omega)$ is the total entropy of P relative to μ . We study the Lebesgue decomposition of H with respect to μ .

Keywords: Entropy, information, measure, probability.

1. Introduction

The concept of “the entropy of a probability measure P ” is a misnomer: there is always another measure μ in the background. In the discrete case considered by Shannon, μ is the cardinality measure [8, p.19]; in the continuous case considered by both Shannon and Wiener, μ is the Lebesgue measure [8, p.54] and [10, pp. 61, 62]. *All entropies are relative entropies of P with respect to some measure μ* , as Shannon and Wiener both emphasized, cf. [8, pp. 57, 58] and [10, pp. 61, 62].

The important paper by Gelfand, Kolmogorov and Yaglom [3] called attention to the case in which the underlying μ is an arbitrary probability measure. This case was studied independently (ca. 1960) by Kallianpur [4] in the United States and Pinsker [6] in the Soviet Union, and perhaps others guided by the earlier work of Kullback and Leibler [5]. They took countably additive (CA) probability measures P, μ on a σ -algebra $\mathfrak{A}$ over an arbitrary set Ω , and taking

Correspondence to: Prof. P.R. Masani, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260, United States.

Π_Ω to be the class of all partitions π of Ω into a finite number of disjoint ($\parallel$) cells Δ in $\mathfrak{A}$, they defined $H_\pi(P|\mu)$ and $H(P|\mu)$, by

$$\begin{aligned} \forall \pi \in \Pi_\Omega, \quad H_\pi(P|\mu) &:= \sum_{\Delta \in \pi+} \left\{ \log \frac{P(\Delta)}{\mu(\Delta)} \right\} \cdot P(\Delta), \\ H(P|\mu) &:= \sup_{\pi \in \Pi_\Omega} H_\pi(P|\mu). \end{aligned} \quad (1.1)$$

Here, $\pi+ = \{\Delta: \Delta \in \pi \text{ \& } \mu(\Delta) > 0 \text{ or } P(\Delta) > 0\}$. The number $H(P|\mu)$ is called *the total entropy of P relative to μ* . A basic result is

$$H(P|\mu) \in [0, \infty]. \quad (1.2)$$

However, *the underlying measure μ cannot always be taken to be a probability measure*. Both Shannon and Wiener took the entropy of a probability P over $\mathbb{R}$ relative to Lebesgue measure μ over $\mathbb{R}$. Indeed, the probability imposition on μ excludes even the elementary case considered by Shannon, as the following easily provided proposition shows.

Proposition 1.1. *Let Ω be finite with two or more members, $\mathfrak{A} = 2^\Omega$, and $\forall \omega \in \Omega, P\{\omega\} > 0$, and let S be the Shannon entropy defined by*

$$S = \sum_{\omega \in \Omega} (\log P\{\omega\}) \cdot P\{\omega\}.$$

Then for all probability measures μ on $\mathfrak{A}$, $S < H(P|\mu)$.

We leave the easy proof to the reader. In the very simplest “uniform” situation where Ω has n members and $\mu\{\omega\} = 1/n, \forall \omega \in \Omega$, we have

$$H(P|\mu) = S + \log n > S, \quad \text{for } n > 1. \quad (1.3)$$

The inequalities in Proposition 1.1 and (1.3) can be turned into equalities by lifting the probability restraint on μ . In Proposition 1.1, for instance, if μ is replaced by the cardinality measure μ_0 on $\Omega = \{1, \dots, n\}$, so that each $\mu_0[k] = 1$, we obviously have $S = H(P|\mu_0)$.¹ If $\Omega = \mathbb{N}_+$, the set of all positive integers, the required measure μ_0 will, like Lebesgue measure, be $[0, \infty]$ -valued.

The definitions of $H_\pi(P|\mu)$, $H(P|\mu)$ in (1.1) break down when μ is not a bounded measure on $\mathfrak{A}$. One purpose of this paper is to provide a universal definition of $H(P|\mu)$ valid for all $[0, \infty]$ -valued measures μ . The few who have dealt with $[0, \infty]$ -valued μ , for instance, Rosenblatt-Roth [7] and Csiszar [1],² have assumed that μ is σ -finite, and have tailored their definitions accordingly. This assumption is satisfied in most applications, but is theoretically unnecessary. The definition we shall give is quite different from theirs, but nevertheless subsumes theirs.

A second purpose of this paper is to show that the subject admits considerable widening when we interpret $H(P|\mu)$ not as a *number* depending on P and μ , but rather as a (P, μ) -dependent *set-function* on the δ -ring $\mathfrak{A}_\mu$ of subsets A of $\mathfrak{A}$ for which $\mu(A) < \infty$ (cf.

¹ This follows on noting that the partition $\pi_0 = \{\{1\}, \dots, \{n\}\}$ has no proper refinements, that $H_\pi(P|\mu)$ increases monotonically with the refinement of π , and that $S = H_{\pi_0}(P|\mu)$.

² The writer is grateful to Prof. A. Barron for this reference.

Definition 2.2). The entropy measure $H := H(P|\mu)$ so obtained turns out to be countably additive on the δ -ring $\mathfrak{A}_\mu$ to $(-\infty, \infty]$ (Theorem 2.9).

It turns out that the Lebesgue decomposition of P into parts P_a and P_b that are absolutely continuous and singular with respect to μ yields a corresponding decomposition of the entropy measure H on $\mathfrak{A}_\mu$ into parts H_a, H_b that are absolutely continuous and singular with respect to μ (Theorems 2.14 and 2.15). Moreover, the classical theorems, due to Kallianpur [4] and others, on the total entropy of P relative to a probability measure μ (which in our notation would read $H(P|\mu)(\Omega)$) admit, for arbitrary μ , full-fledged extensions to $H(P|\mu)(E)$ when $E \in \mathfrak{A}_\mu$, and to $H(P|\mu)(\Omega)$ when $\Omega \in \mathfrak{A}_\mu$, i.e., when μ is bounded, cf. Theorems 2.14 and 2.15.

In Part II (to be published in this journal) we shall show how by taking the total variation $|H|$ of the relative entropy measure H and its Hahn–Jordan decomposition, we can extend H to a CA measure $\bar{H}$ on a larger family $\mathcal{P}$. We will also show that when P has a density f with respect to μ , then $\bar{H}$ itself has a density ϕ of f with respect to μ where $\phi(x) := x \log x$, $x > 0$. We shall also study the invariance of the entropy measure under μ -measure preserving transformations of the space Ω .

2. The relative entropy measure on the fundamental δ -ring

In the rest of this paper we shall adhere to the following notation.

Notation 2.1. (a) $\mathfrak{A}$ is a σ -algebra over Ω ; P is a CA probability or (occasionally) subprobability measure on $\mathfrak{A}$; μ is a CA measure on $\mathfrak{A}$ to $[0, \infty]$;

$$\mathfrak{A}_\mu := \{A: A \in \mathfrak{A} \text{ \& } \mu(A) < \infty\};$$

$$\mathfrak{A}_\mu + := \{A: A \in \mathfrak{A}_\mu \text{ \& } \mu(A) > 0 \text{ or } P(A) > 0\}.$$

(b) $\forall A \in \mathfrak{A}_\mu$,

$$\Pi_A := \{\pi: \pi \text{ is a partition of } A \text{ into a finite number of } \parallel \text{ sets } \Delta \in \mathfrak{A}_\mu\}.$$

For $\pi', \pi \in \Pi_A$, we say that π' is a *refinement* of π , and write $\pi' < \pi$, iff

$$\forall \Delta' \in \pi', \exists \Delta \in \pi \ni \Delta' \subseteq \Delta.$$

(c) $\forall A \in \mathfrak{A}_\mu$ and $\forall \pi \in \Pi_A$,

$$\pi^+ := \{\Delta: \Delta \in \pi \text{ \& } P(\Delta) > 0\}, \quad \pi_+ := \{\Delta: \Delta \in \pi \text{ \& } \mu(\Delta) > 0\}, \quad \pi + := \pi^+ \cup \pi_-.$$

(d) $\forall \Delta \in \pi +$, $\rho(\Delta) := \frac{P(\Delta)}{\mu(\Delta)}$, $Q(\Delta) = \{\log \rho(\Delta)\} \cdot P(\Delta)$.

We note that for $\Delta \in \pi \setminus \pi +$, $P(\Delta) = 0 = \mu(\Delta)$, and $\rho(\Delta) = 0/0$ is meaningless. Moreover, since P and μ are FA, and the function $\phi: \phi(x) = x \log x$, $x \in \mathbb{R}_+$, has $-1/e$ as its absolute minimum, cf. (A.2), we have

$$\forall A \in \mathfrak{A}_\mu + \text{ \& } \forall \pi \in \Pi_A, \pi + \text{ is nonvoid,}$$

$$P(A) = \sum_{\Delta \in \pi^+} P(\Delta) \text{ \& } \mu(A) = \sum_{\Delta \in \pi_+} \mu(\Delta), \quad (2.1)$$

$$\forall \Delta \in \pi +, \quad \rho(\Delta) \in [0, \infty] \text{ \& } -\frac{1}{e} \mu(A) \leq -\frac{1}{e} \mu(\Delta) \leq Q(\Delta) \leq \infty.$$

Likewise, for $A \in \mathfrak{A} \setminus \mathfrak{A}_\mu +$, both π^+ and π_+ are void and so therefore is $\pi +$. These facts suggest the following definition.

Definition 2.2. Let $A \in \mathfrak{A}_\mu$. Then

$$(a) \quad \forall \pi \in \Pi_A, H_\pi(A) = \begin{cases} \sum_{\Delta \in \pi^+} Q(\Delta), & A \in \mathfrak{A}_\mu +, \\ 0, & A \in \mathfrak{A}_\mu \setminus \mathfrak{A}_\mu +; \end{cases}$$

$$(b)^3 \quad H(A) = \sup_{\pi \in \Pi_A} H_\pi(A).$$

An easy consequence of (2.1) and this definition is the following useful triviality.

Triviality 2.3.

- (a) $\forall A \in \mathfrak{A}_\mu, -\frac{1}{e}\mu(A) \leq H(A) \leq \infty,$
 $P(A) = 0 \Rightarrow H(A) = 0, \quad \mu(A) = 0 < P(A) \Rightarrow H(A) = \infty,$
- (b) $\forall A \in \mathfrak{A}_\mu +, -\frac{1}{e}\mu(A) \leq Q(A) \leq H(A) \leq \infty,$
- (c) $\forall A, B \in \mathfrak{A}_\mu, H(A) = \infty \ \& \ A \subseteq B \Rightarrow H(B) = \infty,$
- (d) $\forall A \in \mathfrak{A}_\mu, H(A) < \infty \Rightarrow P \prec_A \mu,$

where $P \prec_A \omega$ means $\Delta \in \mathfrak{A}_\mu \cap 2^A$ & $\mu(\Delta) = 0 \Rightarrow P(\Delta) = 0$.

Proof. (a), (b) These follow readily from (2.1) and Definition 2.2. The trivial case $A \in \mathfrak{A}_\mu \setminus \mathfrak{A}_\mu +$ has to be treated separately.

(c) To prove (c) we associate with each π in Π_A , the partition $\pi^* := \pi \cup \{B \setminus A\} \in \Pi_E$, and we then take the supremum.

(d) Let $A \in \mathfrak{A}_\mu, H(A) < \infty, \Delta \in \mathfrak{A}_\mu \cap 2^A$ and $\mu(A) = 0$. An application of Triviality 2.3(c) with A, B replaced by Δ and A shows that $H(\Delta) < \infty$, whence by Triviality 2.3(b), $Q(\Delta) < \infty$. But for $P(\Delta) > 0$, we have $Q(\Delta) = \infty$; hence $P(\Delta) = 0$. Thus, $P \prec_A \mu$. $\square$

We shall now show that the definition of entropy given in Definition 2.2 is unaffected by the replacement of finite partitions of A by countable partitions of A . Let

$$\forall A \in \mathfrak{A}_\mu, \bar{\Pi}_A := \{\pi: \pi \text{ is a partition of } A \text{ into a finite or countable number of cells in } \mathfrak{A}_\mu\}. \quad (2.2)$$

³ The full symbol for $H(A)$ should be $\{H(P|\mu)\}(A)$ or $H_{P|\mu}(A)$. But since the measures P and μ are fixed, the abbreviation $H(A)$ suffices.

For $A \in \mathfrak{A}_\mu +$, and $\pi \in \bar{\Pi}_A$, the series $\sum_{\Delta \in \pi+} Q(\Delta)$ is infinite, and the question of its unconditional convergence or unconditional divergence arises. But this matter is easily settled: by Triviality 2.3(c) each $Q(\Delta) \geq -e^{-1}\mu(\Delta)$; hence the sum of the negative terms of the series $\sum_{\Delta \in \pi+} Q(\Delta)$ is bounded below by $-e^{-1}\mu(A)$ and cannot be $-\infty$, cf. Lemma A.1. Thus:

Triviality 2.4. $\forall A \in \mathfrak{A}_\mu +$ & $\forall \pi \in \bar{\Pi}_A$, the series

$$H_\pi^*(A) := \sum_{\Delta \in \pi+} Q(\Delta)$$

converges unconditionally to some number in $[-e^{-1}\mu(A), \infty)$, or it diverges unconditionally to $+\infty$.

We now assert the following.

Transition Lemma 2.5. Let $A \in \mathfrak{A}_\mu +$. Then

$$H(A) = \sup_{\pi \in \bar{\Pi}_A} H_\pi^*(A).$$

Proof. Write $H^*(A) := \sup_{\pi \in \bar{\Pi}_A} H_\pi^*(A)$. Since

$$\Pi_A \subseteq \bar{\Pi}_A \quad \& \quad \forall \pi \in \Pi_A, \quad H_\pi^*(A) = H_\pi(A),$$

we have $H(A) \leq H^*(A)$.

To prove the reverse inequality, we may, since $-\infty < -e^{-1}\mu(A) \leq H(A) \leq H^*(A)$, take a number $r_0 \in \mathbb{R} \ni r_0 < \bar{H}(A) \leq \infty$. Then

$$\exists \bar{\pi} \in \bar{\Pi}_A \ni r_0 < H_{\bar{\pi}}^*(A) \leq \bar{H}(A).$$

Let $\bar{\pi} + = \{\Delta_k\}_{k=1}^\infty$, and therefore, cf. Triviality 2.4,

$$r_0 < H_{\bar{\pi}}^*(A) := \sum_{k=1}^\infty Q(\Delta_k).$$

By Triviality 2.4, this series converges unconditionally or diverges to ∞ unconditionally. Hence,

$$(1) \quad \exists n_0 \in \mathbb{N}_+ \ni n \geq n_0 \Rightarrow r_0 < \sum_{k=1}^n Q(\Delta_k).$$

Now obviously $D_{n+1} := \bigcup_{k=n+1}^\infty \Delta_k \in \mathfrak{A}_\mu$, and since $D_{n+1} \supseteq \Delta_{n+1}$ and $\Delta_{n+1} \in \bar{\pi} +$, therefore $D_{n+1} \in \mathfrak{A}_\mu +$. Hence

$$\pi_n := \{\Delta_1, \dots, \Delta_n, D_{n+1}\} \in \Pi_A \quad \& \quad \pi_n + := \pi_n.$$

It follows from Definition 2.2 that

$$H_{\pi_n}(A) := \sum_{k=1}^n Q(\Delta_k) + Q(D_{n+1}) > r_0 - \frac{1}{e}\mu(D_{n+1}), \quad \text{for } n \geq n_0,$$

by (1) and Triviality 2.3(b). Thus, $\forall n \geq n_0$,

$$(2) \quad r_0 < H_{\pi_n}(A) + \frac{1}{e}\mu(D_{n+1}) \leq H(A) + \frac{1}{e}\mu(D_{n+1}).$$

Now, cf. (2.1), $\mu(A) = \sum_{\Delta \in \pi+} \mu(\Delta) = \sum_{k=1}^{\infty} \mu(\Delta_k)$. It follows, since $\mu(A) < \infty$, that

$$\mu(D_{n+1}) = \mu(A) - \sum_{k=1}^n \mu(\Delta_k) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Hence, on letting $n \rightarrow \infty$ in the right-hand side of (2), we get $r_0 \leq H(A)$. Here r_0 is any number less than $\bar{H}(A)$. Thus $H^*(A) \leq H(A)$. $\square$

So far no use has been made of the concavity of the log function on $(0, \infty)$. Its use yields a sharpening of Triviality 2.3(b), and thence the useful result that $H_{\pi}(A)$ and $H_{\pi}^*(A)$ do not diminish under the refinement of π in Π_A follows.

Lemma 2.6.

$$\forall A \in \mathfrak{A}_{\mu+}, \forall \pi \in \Pi_A \text{ \& } \forall \bar{\pi} \in \bar{\Pi}_A, \quad -\frac{1}{e} \mu(A) \leq Q(A) \leq H_{\pi}(A) \text{ \& } Q(A) \leq H_{\bar{\pi}}^*(A).$$

Proof. Let $A \in \mathfrak{A}_{\mu+}$ and $\pi \in \Pi_A$. We have to consider three cases: (1) $P(A) = 0 < \mu(A)$, (2) $\mu(A) = 0 < P(A)$, (3) $P(A) > 0 < \mu(A)$. In case (1), $Q(A) = 0 = H_{\pi}(A)$ and in case (2), $Q(A) = \infty = H_{\pi}(A)$. Thus, the first inequality holds trivially in cases (1) and (2).

In case (3), $\rho(A) := P(A)/\mu(A) \in \mathbb{R}_+$. We may therefore apply the inequality (cf. Lemma A.3):

$$\forall x \in [0, \infty] \text{ \& } \forall x_0 \in \mathbb{R}_+, \quad \log x \geq \log x_0 - \frac{x_0}{x} + 1,$$

with $x = \rho(\Delta)$ and $x_0 = \rho(A)$. Thus

$$\log \rho(\Delta) \geq \log \rho(A) - \frac{\rho(A)}{\rho(\Delta)} + 1.$$

Multiplying the terms by $P(\Delta)$ and summing over Δ in $\pi+$, where $\pi \in \Pi_A$, it follows easily that $\sum_{\Delta \in \pi+} Q(\Delta) \geq Q(A)$, i.e., that $H_{\pi}(A) \geq Q(A)$.

Next, let $\bar{\pi} \in \bar{\Pi}_A$. Then the argument given in the last two paragraphs carries over verbatim with the replacement of π and $H_{\pi}(A)$ by $\bar{\pi}$ and $H_{\bar{\pi}}^*(A)$. $\square$

We need the following simple result on refinements.

Triviality 2.7. Let $A \in \mathfrak{A}_{\mu+}$, $\pi, \pi' \in \Pi_A$ & $\pi' < \pi$. Then

$$\forall \Delta \in \pi, \quad \pi' \cap 2^{\Delta} \in \Pi_A \text{ \& } \pi' + = \sum_{\Delta \in \pi+} (\pi' \cap 2^{\Delta}) +.$$

Proof. Let $\Delta \in \pi$. Then since $\pi' < \pi$, obviously $\pi' \cap 2^{\Delta}$ is a partition of Δ , i.e., $\pi' \cap 2^{\Delta} \in \Pi_A$.

Next let $\Delta' \in \pi' +$ and let $\Delta' \subseteq \Delta \in \pi$. Then since $\mu(\Delta') > 0$ or $P(\Delta') > 0$, it follows that $\mu(\Delta) > 0$ or $P(\Delta) > 0$, i.e., $\Delta \in \pi+$. Thus

$$\Delta' \in \pi' + \Rightarrow \Delta' \in \bigcup_{\Delta \in \pi+} 2^{\Delta},$$

i.e., $\pi' + \subseteq \bigcup_{\Delta \in \pi+} 2^{\Delta}$. Hence $\pi' + = \bigcup_{\Delta \in \pi+} \{(\pi' +) \cap 2^{\Delta}\}$. Since obviously $(\pi' +) \cap 2^{\Delta} = (\pi' \cap 2^{\Delta}) +$, we are done. $\square$

Lemma 2.8. *Let $A \in \mathfrak{A}_{\mu} +$. Then*

$$\pi, \pi' \in \Pi_A \text{ \& } \pi' < \pi \Rightarrow H_{\pi}(A) \leq H_{\pi'}(A),$$

$$\bar{\pi}, \bar{\pi}' \in \bar{\Pi}_A \text{ \& } \bar{\pi}' < \bar{\pi} \Rightarrow H_{\bar{\pi}}^*(A) \leq H_{\bar{\pi}'}^*(A).$$

Proof. Let $\pi, \pi' \in \Pi_A$ and $\pi' < \pi$. Then from the equality $\pi' + = \bigcup_{\Delta \in \pi+} (\pi' \cap 2^{\Delta}) +$, cf. Triviality 2.7, we conclude that

$$\begin{aligned} H_{\pi'}(A) &= \sum_{\Delta' \in \pi'+} Q(\Delta') = \sum_{\Delta \in \pi+} \sum_{\Delta' \in (\pi' \cap 2^{\Delta})+} Q(\Delta') \\ (1) \quad &= \sum_{\Delta \in \pi+} H_{(\pi' \cap 2^{\Delta})}(\Delta), \text{ since } \pi' \cap 2^{\Delta} \in \Pi_{\Delta}. \end{aligned}$$

But by Lemma 2.6, $\forall \Delta \in \pi +$, $Q(\Delta) \leq H_{(\pi' \cap 2^{\Delta})}(\Delta)$. Thus from (1) we get

$$H_{\pi'}(A) \geq \sum_{\Delta \in \pi+} Q(\Delta) =: H_{\pi}(A).$$

Next let $\bar{\pi}, \bar{\pi}' \in \bar{\Pi}_A$ and $\bar{\pi}' < \bar{\pi}$. Then the argument just given carries over with the replacement of $\pi, \pi', H_{\pi}(A), H_{\pi'}(A), H_{(\pi' \cap 2^{\Delta})}(\Delta)$ by $\bar{\pi}, \bar{\pi}', H_{\bar{\pi}}^*(A), H_{\bar{\pi}'}^*(A), H_{(\bar{\pi}' \cap 2^{\Delta})}^*(\Delta)$, respectively. $\square$

An immediate corollary is the result that $H(A)$ is the refinement limit of $H_{\pi}(A)$ and of $H_{\bar{\pi}}^*(A)$:

$$\forall A \in \mathfrak{A}_{\mu} +, \lim_{\substack{\pi \downarrow \\ \pi \in \Pi_A}} H_{\pi}(A) = H(A) = \lim_{\substack{\pi \downarrow \\ \pi \in \bar{\Pi}_A}} H_{\pi}^*(A). \quad (2.3)$$

These results culminate in the following theorem.

Theorem 2.9. *$H(\cdot)$ is a CA measure on the δ -ring $\mathfrak{A}_{\mu}$ to $(-\infty, \infty]$.*

Proof. Let $\forall k \geq 1, A_k \in \mathfrak{A}_{\mu}$, A_k be || & $A = \bigcup_{k=1}^{\infty} A_k \in \mathfrak{A}_{\mu}$. We have to show that $H(A) = \sum_{j=1}^{\infty} H(A_j)$. Since the terms are uniformly bounded below by $-e^{-1}\mu(A) > -\infty$, the series unconditionally converges, or unconditionally diverges to $+\infty$.

Now if A is in $\mathfrak{A}_{\mu} \setminus \mathfrak{A}_{\mu} +$, so is each A_k . By Definition 2.2, $H(A) = 0 = H(A_k)$, and the equality holds trivially.

Let $A \in \mathfrak{A}_{\mu} +$ and let

$$J = \{j: j \in \mathbb{N}_+ \text{ \& } A_j \in \mathfrak{A}_{\mu} +\}.$$

Since, cf. Definition 2.2, for $k \in \mathbb{N}_+ \setminus J$, $H(A_k) = 0$, we have to show that $H(A) = \sum_{j \in J} H(A_j)$. We shall proceed in two steps:

$$(I) \quad H(A) \leq \sum_{j \in J} H(A_j),$$

$$(II) \quad \sum_{j \in J} H(A_j) \leq H(A).$$

Proof of (I). Let $\pi \in \Pi_A$. Then

$$\pi_k := \{\Delta \cap A_k : \Delta \in \pi\} \in \Pi_{A_k}, \quad \pi' = \bigcup_{k=1}^{\infty} \pi_k \in \bar{\Pi}_A \quad \& \quad \pi' < \pi.$$

Obviously $\pi' + = \bigcup_{k=1}^{\infty} \pi_k +$. But since $\forall k \in \mathbb{N}_+ \setminus J$, $\pi_k + = \emptyset$, we have

$$(1) \quad \pi' + = \bigcup_{j \in J} \pi_j +.$$

It follows from (1) that

$$H_{\pi'}^*(A) := \sum_{\Delta \in \pi' +} Q(\Delta) = \sum_{j \in J} \sum_{\Delta \in \pi_j +} Q(\Delta) = \sum_{j \in J} H_{\pi_j}(A_j) \leq \sum_{j \in J} H(A_j).$$

Since $\pi' < \pi$, we conclude from Lemma 2.8 that

$$H_{\pi}(A) = H_{\pi'}^*(A) \leq H_{\pi'}^*(A) \leq \sum_{j \in J} H(A_j).$$

Taking the sup over $\pi \in \Pi_A$, we have (I).

Proof of (II). Let $\forall k \geq 1$, $\pi_k \in \Pi_{A_k}$. Then obviously

$$\pi := \bigcup_{k=1}^{\infty} \pi_k \in \bar{\Pi}_A,$$

and exactly as with (1) above, we have

$$(2) \quad \pi + = \bigcup_{j \in J} \pi_j +.$$

Since $\forall j \in J$, $\pi_j + \neq \emptyset$, we see, cf. Definition 2.2,

$$\sum_{j \in J} H_{\pi_j}(A_j) = \sum_{j \in J} \sum_{\Delta \in \pi_j +} Q(\Delta).$$

By virtue of (2), the Δ appearing on the right-hand side are precisely the cells of $\pi +$. Thus the last summation is a derangement of $\sum_{\Delta \in \pi +} Q(\Delta) = H_{\pi}^*(A)$. Thus

$$(3) \quad \sum_{j \in J} H_{\pi_j}(A_j) = H_{\pi}^*(A) \leq H(A).$$

Now (II) obviously holds if $H(A) = \infty$. Let therefore $H(A) < \infty$. Then by Triviality 2.3(c), each $H(A_k) < \infty$. Letting $\epsilon > 0$, it follows that

$$\forall j \in J, \exists \pi_j \in \Pi_{A_j} \ni H(A_j) - \frac{\epsilon}{2^j} < H_{\pi_j}(A_j).$$

Summing over $j \in J$, we get, cf. (3),

$$\sum_{j \in J} H(A_j) - \epsilon < \sum_{j \in J} H_{\pi_j}(A_j) \leq H(A).$$

Letting $\epsilon \rightarrow 0+$, we have (II). $\square$

It is natural to ask how the absolute continuity or singularity of P with respect to μ bears on the absolute continuity or singularity of the entropic measure H with respect to μ . For the measures at hand, defined on δ -rings, the relevant notions of global and local absolute

continuity and singularity (symbolized by $\prec\prec$, $\prec\prec^{\text{loc}}$ and $\perp$) are defined in Definition A.6. The symbol $\mathfrak{N}_\mu$ denotes the family of μ -negligible sets, cf. Definition A.5.

First as to absolute continuity, we have the following proposition.

Proposition 2.10. (a) $H \prec\prec^{\text{loc}} \mu \Rightarrow \mathfrak{N}_\mu \subseteq \mathfrak{N}_H \Leftrightarrow P \prec\prec \mu$.
 (b) $\text{Range } H \subseteq \mathbb{R} \Rightarrow P \prec\prec \mu \text{ \& } H \prec\prec \mu$.

Proof. (a) The first implication follows from (A.7)(a). Next let $\mathfrak{N}_\mu \subseteq \mathfrak{N}_H$, and let $A \in \mathfrak{A}_\mu$ & $\mu(A) = 0$. Then $H(A) = 0$. Since, cf. Triviality 2.3(b), $-e^{-1}\mu(A) \leq Q(A) \leq H(A)$, it follows that $Q(A) = 0$. If $P(A) > 0$, $Q(A) := \log\{P(A)/0\}P(A) = \infty$. Hence $P(A) = 0$. Thus $\mathfrak{N}_\mu \subseteq \mathfrak{N}_P$. But P and μ satisfy the conditions of Corollary A.8 with $\mathfrak{D} = \mathfrak{A}$, and therefore

$$(1) \quad \mathfrak{N}_\mu \subseteq \mathfrak{N}_P \Leftrightarrow P \prec\prec \mu.$$

Thus $P \prec\prec \mu$.

Finally let $P \prec\prec \mu$. Then by (1), $\mathfrak{N}_\mu \subseteq \mathfrak{N}_P$. Hence

$$\begin{aligned} A \in \mathfrak{A} \text{ \& } \mu(A) = 0 &\Rightarrow A \in \mathfrak{A}_\mu \text{ \& } \mu(A) = 0 = P(A) \\ &\Rightarrow \forall \Delta \in \mathfrak{A}_\mu \cap 2^A, \Delta \in \mathfrak{A}_\mu \setminus \mathfrak{A}_\mu + \\ &\Rightarrow \forall \Delta \in \mathfrak{A}_\mu \cap 2^A, H(\Delta) := 0, \text{ by Definition 2.2,} \\ &\Rightarrow A \in \mathfrak{N}_H. \end{aligned}$$

Thus $\mathfrak{N}_\mu \subseteq \mathfrak{N}_H$. This establishes (a).

(b) Let $\text{Range } H \subseteq \mathbb{R}$ and suppose that $P \nprec\prec \mu$. Then by (1), $\mathfrak{N}_\mu \not\subseteq \mathfrak{N}_P$, i.e.,

$$(2) \quad \exists A_0 \in \mathfrak{N}_\mu \ni \mu(A_0) = 0 < P(A_0).$$

But since $H(A_0) < \infty$, Triviality 2.3(d) tells us that $P \prec\prec_{A_0} \mu$, i.e.,

$$(3) \quad \Delta \in \mathfrak{N}_\mu \cap 2^{A_0} \text{ \& } \mu(\Delta) = 0 \Rightarrow P(\Delta) = 0.$$

Taking $\Delta = A_0$ in (3), it follows, since $\mu(A_0) = 0$, that $P(A_0) = 0$. This contradicts the inequality in (2), and shows that our supposition is untenable; thus $P \prec\prec \mu$.

To show that $H \prec\prec \mu$, we have, as just shown, $P \prec\prec \mu$. Hence by (a),

$$(4) \quad \mathfrak{N}_\mu \subseteq \mathfrak{N}_H.$$

But since $\text{Range } H \subseteq \mathbb{R}$, H is a CA measure on $\mathfrak{A}_\mu$ to $\mathbb{R}$. Hence, as is well known,

$$\forall A \in \mathfrak{A}_\mu, |H|(A) < \infty,$$

where $|H|$ is the total variation of H , cf. (A.4). The measures μ and H therefore satisfy the conditions of Corollary A.8 on $\mathfrak{D} = \mathfrak{A}_\mu$. Hence by Corollary A.8, $\mathfrak{N}_\mu \subseteq \mathfrak{N}_H \Rightarrow H \prec\prec \mu$. This together with (4) yields $H \prec\prec \mu$.

This completes the proof of (b). $\square$

To deal with the issue of singularity, we need the following triviality.

Triviality 2.11. Let $C \in \mathfrak{A}$. Then, cf. Definition A.5(a),
 C is a carrier of $P \Rightarrow C$ is a carrier of H .

Proof. Let C be a carrier of P . Then $\forall A \in \mathfrak{A}_\mu$, $P(A \setminus C) = 0$, whence by Triviality 2.3(a), $H(A \setminus C) = 0$. Thus C is a carrier of H . $\square$

Proposition 2.12. $P \perp \mu \Rightarrow H \perp \mu \Rightarrow \text{Range } H = \{0, \infty\}$.

Proof. Let $P \perp \mu$. Then P and μ have disjoint carriers C_P, C_μ . Since by Triviality 2.11, C_P is a carrier of H , we have $H \perp \mu$.

Next let $H \perp \mu$, and let C_H, C_μ be their disjoint carriers. Let $A \in \mathfrak{A}_\mu$ and $H(A) \neq 0$. To finish the proof, we only have to show that

$$(I) \quad H(A) = \infty.$$

Proof of (I). Since $A \cap C_H \parallel C_\mu$, therefore $\mu(A \cap C_H) = 0$. But since $H(A \cap C_H) = H(A) \neq 0$, therefore, by Definition 2.2, $A \cap C_H \in \mathfrak{A}_\mu +$. Thus $P(A \cap C_H) > 0$. Hence by Triviality 2.3(a), $H(A \cap C_H) = \infty$. Since $H(A \cap C_H) = H(A)$, we have (I). $\square$

The next example shows that the implication $P \perp \mu \Rightarrow \text{Range } H = \{0, \infty\}$ in Proposition 2.12 is *not* reversible in general.

Example 2.13. Let $\Omega = (0, 1]$, $\mathfrak{A}$ be the σ -algebra of Borel subsets of $(0, 1]$, μ be the Lebesgue measure on $\mathfrak{A}$ and

$$\forall A \in \mathfrak{A}, \quad P(A) := \mu\{A \cap (0, \tfrac{1}{2}]\} + \tfrac{1}{2}\chi_A(1).$$

Then $\text{Range } H = \{0, \infty\}$, but $P \not\perp \mu$.

Proof. Writing $\Omega_0 = (0, \tfrac{1}{2}]$, $\Omega_1 = (\tfrac{1}{2}, 1)$, we have

$$(1) \quad \forall A \in \mathfrak{A}, \quad H(A) = H(A \cap \Omega_0) + H(A \cap \Omega_1) + H(A \cap \{1\}).$$

But $H(A \cap \Omega_0) = 0$, since for $\Delta \subseteq A \cap \Omega_0$, $P(\Delta) := \mu(\Delta)$ and therefore $\mathcal{C}(\Delta) = 0$. Next, since $P(A \cap \Omega_1) := 0$, therefore by Triviality 2.3(a), $H(A \cap \Omega_1) = 0$. Finally, since $\mu\{1\} = 0 < P\{1\}$, by Triviality 2.3(a), $H\{1\} = \infty$. Thus (1) reduces to

$$\forall A \in \mathfrak{A}, \quad H(A) = \infty \cdot \chi_A\{1\},$$

and $\text{Range } H = \{0, \infty\}$.

Let C_P be a carrier of P . Then $1 \in C_P$ and

$$1 = P(C_P) = P(C_P \cap \Omega_0) + P\{1\} = \mu(C_P \cap \Omega_0) + \tfrac{1}{2},$$

and therefore $\mu(C_P) \geq \mu(C_P \cap \Omega_0) = \tfrac{1}{2}$. Let C_μ be a carrier of μ . Were $C_\mu \perp C_P$, then, since μ is FA,

$$\mu(\Omega) \geq \mu(C_\mu) + \mu(C_P) \geq 1 + \tfrac{1}{2},$$

which is absurd. Thus $C_\mu \not\perp C_P$. Hence $P \not\perp \mu$. $\square$

As stated in Remark A.11, the probability P has a Lebesgue decomposition with subprobability components P_a, P_b that are absolutely continuous and singular with respect to μ . In the next theorem we show that the entropy measures of these components yield the corresponding

components in the Lebesgue decomposition of H itself with respect to μ . Briefly, there is a concordance between the two Lebesgue decompositions.

Concordance Theorem 2.14 (Form I). *Let, cf. Theorem A.10 and Remark A.11,*

(i) Ω_b be such that $\Omega_b \in \mathfrak{N}_\mu$ & $\forall N \in \mathfrak{N}_\mu, P(N \setminus \Omega_b) = 0$,

$$\Omega_a := \Omega \setminus \Omega_b, \quad P_a(\cdot) := P(\cdot \cap \Omega_a), \quad P_b(\cdot) := P(\cdot \cap \Omega_b), \quad \text{on } \mathfrak{A},$$

so that $P = P_a + P_b$, $P_a \prec \mu$ & $\mu \perp P_b \perp P_a$;

(ii) H, H_a, H_b be the entropy measures on $\mathfrak{A}_\mu$ of P, P_a, P_b relative to μ , respectively. Then

$$(a) \quad H_a(\cdot) = H(\cdot \cap \Omega_a), \quad H_b(\cdot) = H(\cdot \cap \Omega_b), \quad \text{on } \mathfrak{A}_\mu;$$

$$(b) \quad H = H_a + H_b, \quad H_a \prec^{\text{loc}} \mu \quad \& \quad \mu \perp H_b \perp H_a.$$

Proof. (a) Let $A \in \mathfrak{A}_\mu$. Since, cf. (i), Ω_a is a carrier of P_a , therefore by Triviality 2.11 it is a carrier of H_a . Hence

$$(1) \quad H_a(A) = H_a(A \cap \Omega_a).$$

Now write $\forall \pi \in \Pi_A$,

$$\pi_a + := \{\Delta: \Delta \in \pi \text{ & } \Delta(\Delta) > 0 \text{ or } P_a(\Delta) > 0\}$$

and

$$\forall \Delta \in \pi_a +, \quad Q_a(\Delta) := \log \frac{P_a(\Delta)}{\mu(\Delta)} \cdot P_a(\Delta),$$

and let $A \in \mathfrak{A}_\mu \cap 2^{\Omega_a}$. Since, cf. (i), $P_a = P$ over Ω_a , it follows at once that

$$\forall \pi \in \Pi_A, \quad \pi_a + = \pi + \quad \& \quad \Delta \in \pi_a + \Rightarrow Q_a(\Delta) = Q(\Delta).$$

Hence

$$H_a(A) := \sup_{\pi \in \Pi_A} \sum_{\Delta \in \pi_a +} Q_a(\Delta) = \sup_{\pi \in \Pi_A} \sum_{\Delta \in \pi +} Q(\Delta) = H(A).$$

Thus

$$(2) \quad \forall A \in \mathfrak{A}_\mu \cap 2^{\Omega_a}, \quad H_a(A) = H(A).$$

Combining (1) and (2), we see that

$$\forall A \in \mathfrak{A}_\mu, \quad H_a(A) = H_a(A \cap \Omega_a) = H(A \cap \Omega_a).$$

In exactly the same way we can show that

$$\forall A \in \mathfrak{A}_\mu, \quad H_b(A) = H(A \cap \Omega_b).$$

Thus (a) follows.

(b) Since $\Omega_a \cap \Omega_b = \Omega$ & $\Omega_a \parallel \Omega_b$, and H is FA on $\mathfrak{A}_\mu$, it follows from (a) that

$$(3) \quad H(\cdot) = H(\cdot \cap \Omega_a) + H(\cdot \cap \Omega_b) = H_a(\cdot) + H_b(\cdot), \quad \text{on } \mathfrak{A}_\mu.$$

Next since $P_a \ll \mu$, therefore by (ii) and Proposition 2.10(a),

$$(4) \quad \mathfrak{N}_\mu \subseteq \mathfrak{N}_{H_a}, \quad \text{i.e., } H_a \ll^{\text{loc}} \mu.$$

Finally, since $P_b \perp \mu$, therefore by Proposition 2.12, $H_b \perp \mu$. Also by (a), H_a, H_b have disjoint carriers. Thus

$$(5) \quad \mu \perp H_b \perp H_a.$$

By (3)–(5), we have (b). $\square$

The concordance is also stateable in terms of the Lebesgue decomposition P into probability components, but in a less elegant form. This result is easily deduced from Theorem 2.14.

Concordance Theorem 2.15 (Form II). *Let*

(i) *be as in Theorem 2.14(i);*

(ii) $\lambda_a := P(\Omega_a), \lambda_b := P(\Omega_b), P^a := \lambda_a^{-1}P_a$ and $P^b := \lambda_b^{-1}P_b$ on $\mathfrak{A}$, so that P^a, P^b are the absolutely continuous and singular probability components of P with respect to μ on $\mathfrak{A}$;

(ii) H^a, H^b be the entropy measures on $\mathfrak{A}_\mu$ of the probability measures P^a, P^b relative to μ ;

(iv) $H_a := H^a - \phi(1/\lambda_a)P_a, H_b := H^b - \phi(1/\lambda_b)P_b$ on $\mathfrak{A}_\mu$, where $\phi(x) := x \log x, x > 0$.

Then $H = \lambda_a H_a + \lambda_b H_b, H_a \ll^{\text{loc}} \mu$ & $\mu \perp H_b \perp H_a$.

Appendix A

In this Appendix we discuss some results ancillary to this part of the paper.

In the theory of relative entropy measure, the following result on the unconditional convergence or divergence of a series $\sum u_n$ of terms in $(-\infty, \infty]$ for which $\sum_{k=1}^\infty |u_k| = \infty$ is required.

Lemma A.1. *Let $(u_n)_{n=1}^\infty$ be a sequence in $(-\infty, \infty]$ such that $-\infty < -a_n \leq u_n$, where $a_n \geq 0$ & $\sum_{n=1}^\infty a_n < \infty$. Then the series $\sum u_n$ either converges unconditionally to some number in $\mathbb{R}$ or diverges unconditionally to $+\infty$.*

The proof is quite elementary. Also required is the following somewhat related version of the Fatou Lemma. This involves the concept of a δ -ring of subsets of Ω , i.e., a ring closed under countable intersections, but not necessarily under countable unions. We also need the notation:

$$\text{For } \emptyset \neq \mathcal{F} \subseteq 2^\Omega, \mathcal{F}^{\text{loc}} := \{A: A \subseteq \Omega \text{ & } \forall F \in \mathcal{F}, A \cap F \in \mathcal{F}\}. \quad (\text{A.1})$$

It is easily seen that: for a δ -ring $\mathcal{D}$, $\mathcal{D} \subseteq \sigma\text{-alg}(\mathcal{D}) \subseteq \mathcal{D}^{\text{loc}}$, the last being a σ -algebra. The result in question then reads as follows.

Lemma A.2 (Fatou, for sequences uniformly bounded below). *Let*

- (i) *the CA measure μ on the δ -ring $\mathcal{D}$ with values in $\mathbb{R}_{0+}$ be bounded, i.e., $\sup_{D \in \mathcal{D}} \mu(D) < \infty$;*
- (ii) *the sequence $(f_n)_{n=1}^\infty$ of functions f_n on Ω to $(-\infty, \infty]$ be $\mathcal{D}^{\text{loc}}$, Borel measurable and uniformly bounded below, i.e., $m := \inf_{n \geq 1} f_n \in \mathbb{R}$.*

Then

$$-\infty \leq m \cdot \mu(\Omega) \leq \int_{\Omega} \lim_{n \rightarrow \infty} f_n(\omega) \mu(d\omega) \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n(\omega) \mu(d\omega).$$

This follows readily on applying the classical Fatou Lemma to the sequence of nonnegative functions $f_n(\cdot) - m$.

The concavity of the log function is crucial to the theory, and best adapted to our purpose is the following consequence of this property.

Lemma A.3.

$$\forall x_0 \in \mathbb{R}_+ \ \& \ \forall x \in [0, \infty], \ \log x \geq \log x_0 - \frac{x_0}{x} + 1.$$

Proof. For $x = \infty$, the left-hand side is ∞ , and for $x = 0$, the right-hand side is $-\infty$; hence the inequality holds. Next, let $x \in \mathbb{R}_+$. Then x is an interior point of the domain of the convex function $-\log$. Hence

$$(1) \quad -\log x + \log x_0 \geq \lambda_0(x - x_0),$$

where $\lambda_0 = -\log' x_0 = -1/x_0$. Substituting $\lambda_0 = -1/x_0$ in (1), and writing $y = 1/x$ and $y_0 = 1/x_0$, (1) reduces to

$$\log y \geq \log y_0 - \frac{y_0}{y} + 1.$$

Since $y = 1/x \in \mathbb{R}_{0+}$ & $y_0 = 1/x_0 \in \mathbb{R}_+$, we are done. $\square$

As for the associated function ϕ defined in Part II, Assumption 4.1(b), viz.

$$\phi(x) = x \log x, \quad x \in \mathbb{R}_+,$$

we need only note that

$$\left. \begin{array}{l} \phi \text{ is continuous and convex on } \mathbb{R}_+ \text{ onto } \mathbb{R}_+, \phi(0+) = 0 = \phi(1) \ \& \ \phi(\infty-) = \infty, \\ \phi \text{ is negative on } (0, 1) \text{ and positive on } (1, \infty) \text{ and attains its absolute minimum} \\ -1/e \text{ at } x = 1/e, \\ \forall x, y \geq 0, \ \phi(x \cdot y) = x\phi(y) + y\phi(x). \end{array} \right\} \quad (\text{A.2})$$

The definition of the total variation $|\xi|$ of a CA measure ξ on a δ -ring $\mathcal{D}$ over Ω with values in $\mathfrak{X}$, where ⁴

$$\mathfrak{X} = \text{a normed vector space or } (-\infty, \infty], \quad (\text{A.3})$$

⁴ $\mathfrak{X}$ will have this meaning throughout this Appendix. For purposes of entropy, the normed vector space can be assumed to be $\mathbb{R}$.

is that its domain is the localization σ -algebra $\mathfrak{D}^{\text{loc}}$, cf. (A.1), and

$$\forall E \in \mathfrak{D}^{\text{loc}}, |\xi|(E) := \sup_{\pi \in \Pi} \sum_{\Delta \in \pi} |\xi(E \cap \Delta)|, \text{ where } \Pi := \bigcup_{D \in \mathfrak{D}} \Pi_D. \quad (\text{A.4})$$

It follows readily (cf., e.g., [2, p.35, No.9]) that

$$\left. \begin{array}{l} \text{(a)} \quad |\xi| \text{ is a CA measure on } \mathfrak{D}^{\text{loc}} \text{ to } [0, \infty]; \\ \text{(b)} \quad \forall E \in \mathfrak{D}^{\text{loc}}, |\xi|(E) = \sup_{D \in \mathfrak{D}} |\xi|(E \cap D); \\ \text{(c)} \quad \text{for } [0, \infty]\text{-valued } \xi \text{ on } \mathfrak{D}, \\ \quad \forall E \in \mathfrak{D}^{\text{loc}}, |\xi|(E) = \sup_{D \in \mathfrak{D}} \xi(E \cap D), \text{ and so } \xi \subseteq |\xi|. \end{array} \right\} \quad (\text{A.5})$$

We will need the following simple result; the proof of which is obvious.

Triviality A.4. Let $A \in \mathfrak{D}^{\text{loc}}$ and $\mu_0(\cdot) := \mu(\cdot \cap A_0)$ on $\mathfrak{D}$. Then $\forall E \in \mathfrak{D}^{\text{loc}}$,
 $|\mu_0|(E) = |\mu|(E \cap A_0)$.

Definition A.5. Let $\emptyset \neq \mathfrak{F} \subseteq 2^\Omega$, $\mathfrak{X}$ be as in (A.3), and ξ be a function on $\mathfrak{F}$ to $\mathfrak{X}$. We say that

(a) C is a *carrier* of ξ iff $C \in \mathfrak{F}^{\text{loc}}$ & $\forall F \in \mathfrak{F}$, $\xi(C \cap F) = \xi(F)$;

(b) N is ξ -*negligible* iff $N \in \mathfrak{F}^{\text{loc}}$ & $\forall F \in \mathfrak{F}$, $\xi(N \cap F) = 0$.

$\mathfrak{N}_\xi$ denotes the family of ξ -negligible subsets of Ω .

Definition A.6. Let

(i) $\mathfrak{R}$ be a ring over Ω & $\mathfrak{X}$ be as in (A.3);

(ii) μ and ξ be CA measures on $\mathfrak{R}$ to $[0, \infty]$ and $\mathfrak{X}$, respectively.

We say that

(a) ξ is *absolutely continuous* with respect to μ , and write $\xi \ll \mu$, iff $\forall A \in \mathfrak{R}$ & $\forall \epsilon > 0$, $\exists \delta_{A,\epsilon} > 0$ such that

$$R \in \mathfrak{R} \cap 2^A \text{ & } \mu(R) < \delta_{A,\epsilon} \Rightarrow |\xi(A)| < \epsilon;$$

(b) ξ is *locally absolutely continuous* with respect to μ and write $\xi \ll^{\text{loc}} \mu$,⁵ iff $\forall A \in \mathfrak{R} \ni |\xi|(A) < \infty$ & $\forall \epsilon > 0$, $\exists \delta_{A,\epsilon} > 0$ such that

$$R \in \mathfrak{R} \cap 2^A \text{ & } \mu(R) < \delta_{A,\epsilon} \Rightarrow |\xi(A)| < \epsilon;$$

(c) ξ and μ are *mutually singular*, and write $\xi \perp \mu$, iff there exist disjoint sets $C_1, C_2 \in \mathfrak{R}^{\text{loc}}$ such that C_1 is a carrier of ξ and C_2 is a carrier of μ .

It is trivial that

for a CA measure on $\mathfrak{D}$ to $\mathfrak{X}$,

$$\left. \begin{array}{l} \text{(a)} \quad \mathfrak{N}_\xi = \mathfrak{N}_{|\xi|} \text{ and } \xi \text{ & } |\xi| \text{ have the same carriers;} \\ \text{(b)} \quad C \text{ is a carrier of } \xi \text{ & } C \subseteq \bar{C} \in \mathfrak{D}^{\text{loc}} \Rightarrow \bar{C} \text{ is carrier of } \xi; \\ \text{(c)} \quad C_1, C_2 \text{ are carriers of } \xi \Rightarrow C_1 \cap C_2 \text{ is a carrier of } \xi. \end{array} \right\} \quad (\text{A.6})$$

⁵ We write $\ll^{\text{loc}}$ rather than $\ll_{\text{loc}}$ to stress that “locally” here refers to sets that are “finite” with respect to ξ , not μ .

We readily see that

$$\begin{aligned} (a) \quad \xi \prec \mu &\Rightarrow \xi \prec^{\text{loc}} \mu \Rightarrow \mathfrak{N}_\mu \subseteq \mathfrak{N}_\xi; \\ (b) \quad \mathfrak{N}_\mu \subseteq \mathfrak{N}_\xi &\Leftrightarrow \text{any carrier of } \mu \text{ is a carrier of } \xi. \end{aligned} \quad (\text{A.7})$$

When $\mathfrak{N}$ is a δ -ring, this result improves significantly as can be shown by adapting the proof of von Neumann in [9, p.199, 11.2.4].

Theorem A.7. *Let $\mathfrak{N} = \mathfrak{D}$ be a δ -ring and ξ, μ be as in Definition A.6. Then*

$$\xi \prec \mu \Rightarrow \xi \prec^{\text{loc}} \mu \Leftrightarrow \mathfrak{N}_\mu \subseteq \mathfrak{N}_\xi.$$

Corollary A.8. *Let the measure ξ in Theorem A.7 be such that $\forall D \in \mathfrak{D}, |\xi|(D) < \infty$. Then*

$$\xi \prec \mu \Leftrightarrow \mathfrak{N}_\mu \subseteq \mathfrak{N}_\xi.$$

Proof. For any $\xi \ni \forall D \in \mathfrak{D}, |\xi|(D) < \infty$, it follows from Definition A.6 that $\xi \prec^{\text{loc}} \mu$ iff $\xi \prec \mu$. Hence by Theorem A.7 we have the result. $\square$

For real- or complex-valued CA measures on $\mathfrak{D}$ the condition that $|\xi|(D) < \infty, \forall D \in \mathfrak{D}$, is automatically fulfilled. This is not of course the case for the $(-\infty, \infty]$ -valued entropy measure. (Nor might we add is it the case for most of the significant Banach-spaces-valued measures.)

From (A.6) and Theorem A.7, it follows readily that

if ξ is CA on a δ -ring $\mathfrak{D}$ to $\mathfrak{X}$ and μ is CA on $\mathfrak{D}$ to $[0, \infty]$, then

$$\xi \prec^{\text{loc}} \mu \text{ iff } |\xi| \prec^{\text{loc}} |\mu|, \quad \xi \perp \mu \text{ iff } |\xi| \perp |\mu|. \quad (\text{A.8})$$

To turn to the Lebesgue decomposition, we shall first study the consequences that follow from the assumption of the Lebesgue decomposability of an $\mathfrak{X}$ -valued CA measure on $\mathfrak{D}$ with respect to a $[0, \infty]$ -valued CA measure μ on $\mathfrak{D}$.

Theorem A.9. *Let*

- (i) $\mathfrak{D}$ be a δ -ring over Ω ;
- (ii) μ and ξ be CA measures on $\mathfrak{D}$ to $(0, \infty]$ and $\mathfrak{X}$, respectively;
- (iii) ξ_a, ξ_b be CA measures on $\mathfrak{D}$ to $\mathfrak{X} \ni$

$$\xi = \xi_a + \xi_b, \quad \xi_a \prec^{\text{loc}} \mu \quad \& \quad \mu \perp \xi_b \perp \xi_a.$$

Then

- (a) *the decomposition in (iii) is unique;*
- (b) $\exists \Omega_a, \Omega_b \in \mathfrak{D}^{\text{loc}}$ *such that $\Omega_a \cup \Omega_b = \Omega$, $\Omega_a \parallel \Omega_b$ and Ω_a is a carrier of μ and ξ_a , and Ω_b is a carrier of ξ_b ; moreover,*

$$\xi_a(\cdot) = \xi(\cdot \cap \Omega_a), \quad \xi_b(\cdot) = \xi(\cdot \cap \Omega_b), \quad \text{on } \mathfrak{D};$$

$$(c) \quad |\xi| = |\xi_a| + |\xi_b|, \quad |\xi_a| \prec^{\text{loc}} |\mu| \quad \& \quad |\mu| \perp |\xi_b| \perp |\xi_a|.$$

Proof. (a) The proof is left to the reader.

(b) By (iii), μ and Ω_b have $\parallel$ carriers, say C_1 and C_b . Next, since by (iii), $\mathfrak{N}_\mu \subseteq \mathfrak{N}_{\xi_a}$, therefore by (A.7)(b), μ and ξ have a common carrier, say C_2 . By (A.6)(c),

$$(1) \quad \Omega_a := C_1 \cap C_2 \text{ is a carrier of } \mu \text{ and so of } \nu_a.$$

Next, since $\Omega_a \subseteq C_2 \parallel C_b$, therefore $C_b \subseteq \Omega \setminus \Omega_a \in \mathfrak{D}^{\text{loc}}$. Hence by (A.6)(b),

$$(2) \quad \Omega_b := \Omega \setminus \Omega_a \text{ is a carrier of } \nu_b \text{ \& } \Omega_b \parallel \Omega_a.$$

It follows from (iii), (1) and (2) that $\forall D \in \mathfrak{D}$,

$$\xi(D \cap \Omega_a) = \xi_a(D \cap \Omega_a) + \xi_b(D \cap \Omega_a) = \xi_a(D),$$

and likewise with Ω_b . Thus

$$(3) \quad \xi_a(\cdot) = \xi(\cdot \cap \Omega_a) \text{ \& } \xi_b(\cdot) = \xi(\cdot \cap \Omega_b), \text{ on } \mathfrak{D}.$$

By (1)–(3), we have (b).

(c) It follows from (3) and Triviality A.4 that

$$(4) \quad |\xi_a|(\cdot) = |\xi|(\cdot \cap \Omega_a), \quad |\xi_b|(\cdot) = |\xi|(\cdot \cap \Omega_b), \quad \text{on } \mathfrak{D}^{\text{loc}}.$$

Since $|\xi|$ is FA on $\mathfrak{D}^{\text{loc}}$, we see that $\forall A \in \mathfrak{D}^{\text{loc}}$,

$$(5) \quad |\xi|(A) = |\xi|(A \cap \Omega_a) + |\xi|(A \cap \Omega_b) = |\xi_a|(A) + |\xi_b|(A).$$

Next, since $\mathfrak{N}_\mu \subseteq \mathfrak{N}_{\nu_a}$, therefore by (A.6)(a),

$$(6) \quad \mathfrak{N}_{|\mu|} \subseteq \mathfrak{N}_{|\nu_a|}, \text{ i.e., } |\nu_a| \prec^{\text{loc}} |\mu|.$$

Finally, from (1), (2) and (A.6)(a), Ω_a is a carrier of $|\mu|$ and $|\xi_a|$, and Ω_b is a carrier of $|\nu_b|$. Hence

$$(7) \quad |\mu| \perp |\nu_b| \perp |\nu_b|.$$

By (5)–(7), we have (c). $\square$

Finally, we state a condition on ξ sufficient to ensure that it has a Lebesgue decomposition with respect to μ .

Theorem A.19 (Lebesgue decomposition). *Let*

- (i) $\mathfrak{D}$ be a δ -ring over Ω ;
- (ii) μ and ξ be CA measures on $\mathfrak{D}$ to $[0, \infty]$ and $\mathfrak{X}$, respectively, and $|\xi|(\Omega) < \infty$.

Then

- (a) $\exists \Omega_b \in \mathfrak{N}_\mu \ni \forall N \in \mathfrak{N}_\mu, |\xi|(N \setminus \Omega_b) = 0$;
- (b) letting $\Omega_a := \Omega \setminus \Omega_b$, and letting

$$\forall D \in \mathfrak{D}, \quad \xi_a(D) := \xi(D \cap \Omega_a), \quad \xi_b(D) := \xi(D \cap \Omega_b),$$

we have ξ_a, ξ_b are CA on $\mathfrak{D}$ to $\mathfrak{X}$, and

$$\xi = \xi_a + \xi_b, \quad \xi_a \prec^{\text{loc}} \mu \text{ \& } \mu \perp \xi_b \perp \xi_a;$$

(c) *the decomposition*

$$\xi = \xi_a + \xi_b, \quad \xi_a \prec^{\text{loc}} \mu \quad \& \quad \mu \perp \xi_b \perp \xi_a,$$

where ξ_a, ξ_b are CA on $\mathfrak{D}$ to $\mathfrak{X}$, is unique.

Proof. (a) By (ii) and (A.5)(a),

$$|\mu| \quad \& \quad |\xi| \text{ are CA on } \mathfrak{D}^{\text{loc}} \text{ to } [0, \infty],$$

moreover,

$$\beta := \sup_{N \in \mathfrak{N}_\mu} |\xi|(N) \leq |\xi|(\Omega) < \infty.$$

It follows that $\exists \Omega_b \in \mathfrak{N}_\mu$ such that $|\xi|(\Omega_b) = \beta$. From this it follows in turn that $\forall N \in \mathfrak{N}_\mu$, $|\xi|(N \setminus \Omega_b) = 0$. Thus (a) follows.

(b) Let $\Omega_a, \Omega_b, \xi_a, \xi_b$ be defined as in (a). Then obviously ξ_a, ξ_b are CA on $\mathfrak{D}$ to $\mathfrak{X}$. And since ξ is FA on $\mathfrak{D}$, we have

$$(1) \quad \xi(\cdot) = \xi(\cdot \cap \Omega_a) + \xi(\cdot \cap \Omega_b) = \xi_a(\cdot) + \xi_b(\cdot), \quad \text{on } \mathfrak{D}.$$

Next Ω_a, Ω_b are carriers of ξ_a, ξ_b . Also, since by (a), $\Omega_b \in \mathfrak{N}_\mu$, therefore $\Omega_a := \Omega \setminus \Omega_b$ is a carrier of μ . Since $\Omega_a \parallel \Omega_b$, we see that

$$(2) \quad \xi_a \perp \xi_b \quad \& \quad \mu \perp \xi_b.$$

Finally, let $N \in \mathfrak{N}_\mu$. Then

$$|\xi|(N \cap \Omega_a) = |\xi|(N \setminus \Omega_b) = 0, \quad \text{by (a).}$$

Hence $\forall D \in \mathcal{D}$, $0 = \xi(N \cap \Omega_a \cap D) =: \xi_a(N \cap D)$, i.e., by Definition A.6, $N \in \mathfrak{N}_{\nu_a}$. Thus

$$(3) \quad \mathfrak{N}_\mu \subseteq \mathfrak{N}_{\nu_a}, \quad \text{i.e., } \nu_a \prec^{\text{loc}} \mu.$$

By (1)–(3), we have (b). $\square$

Remark A.11. Let $\mathfrak{A}, P$ and μ be as in Notation 2.1. Then with $\mathfrak{D} = \mathfrak{A}$ and $\nu = P$ the premisses of Theorem A.10 are fulfilled. Furthermore, by Corollary A.8, $\prec^{\text{loc}}$ is replaceable by $\prec$, and $\mathfrak{D}^{\text{loc}} = \mathfrak{D}$. Hence from Theorem A.10, we conclude that

$$\exists \Omega_b \in \mathfrak{N}_\mu \ni \forall N \in \mathfrak{N}_\mu, \quad P(N \setminus \Omega_b) = 0,$$

and letting $\Omega_a := \Omega \setminus \Omega_b$, and

$$P_a(\cdot) := P(\cdot \cap \Omega_a), \quad P_b(\cdot) := P(\cdot \cap \Omega_b), \quad \text{on } \mathfrak{A},$$

we have the Lebesgue decomposition

$$P = P_a + P_b, \quad P_a \prec \mu \quad \& \quad \mu \perp P_b \perp \bar{F}_a.$$

Here P_a, P_b are CA subprobability measures on $\mathfrak{A}$.

References

- [1] I. Csiszar, On generalized entropy, *Studia Sci. Math. Hungar.* **4** (1969) 401–419.
- [2] N. Dinculeanu, *Vector Measures* (Pergamon, Oxford, 1953).
- [3] I.M. Gelfand, N.A. Kolmogorov and A.M. Yaglom, On the general definition of the amount of information, *Dokl. Akad. Nauk USSR* **111** (4) (1956) 745–748 (in Russian).
- [4] G. Kallianpur, On the amount of information contained in a σ -field, in: I. Olkin and S.G. Ghurye, Eds., *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling* (Stanford Univ. Press, Stanford, CA, 1960) 265–273.
- [5] S. Kullback and R.A. Leibler, On information and sufficiency, *Ann. Math. Statist.* **22** (1951) 79–86.
- [6] M.S. Pinsker, *Information and Information Stability of Random Variables and Processes* (Holden-Day, San Francisco, CA, 1964); (Russian ed., 1960).
- [7] M. Rosenblatt-Roth, The concept of entropy in probability theory and its application in the theory of information transmission through communication channels, *Theory Probab. Appl.* **IX** (2) (1964) 212–235.
- [8] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana, IL, 1949).
- [9] J. Von Neumann, *Functional Operators, I* (Princeton Univ. Press, Princeton, NJ, 1950).
- [10] N. Wiener, *Cybernetics* (MIT Press, Cambridge, MA, 2nd ed., 1961).