On the Degree of Mutual Dependence of Three Events

Valentin Vankov Iliev

•

Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Sofia, Bulgaria
viliev@math.bas.bg

"...one of the most important problems
in the philosophy of natural sciences is
... to make precise premises which would
make it possible to regard any given
real events as independent."

A. N. Kolmogorov,
Foundations of the Theory of Probability

Abstract

We define degree of mutual dependence of three events in a probability space by using Boltzmann-Shannon entropy function of an appropriate variable distribution produced by these events and depending on four parameters varying, in general, within of a polytope. It turns out that the entropy function attains its absolute maximum exactly when the three events are mutually independent and its absolute minimum at some vertices of the polytope where the events are "maximally" dependent. By composing the entropy function with an appropriate linear function we obtain a continuous "degree of mutual dependence" function with the same domain and the interval $[0,1]$ *as a target. It attains value* 0 *when the events are mutually independent (the entropy is maximal) and value* 1 *when they are "maximally" dependent (the entropy is minimal). A link is available for downloading a Java code which evaluates the degree of mutual dependence of three events in the classical case of a sample space with equally likely outcomes.*

Keywords: entropy; average information; degree of dependence; probability space; probability distribution; experiment in a sample space; linear system; affine isomorphism; classification space.

1. Introduction

In our papers [6] and [7]) we introduce and study a measure of dependence of two events in a probability space, based on the fundamental notion of Boltzmann-Shannon entropy. The present work is written as a natural conceptual continuation of the above papers for the case of three events A_1, A_2, A_3. By analogy, we consider the joint experiment $\mathfrak{J}_3$ of the corresponding three binary trials, whose probability distribution gives rise to the entropy function that, in turn, measures the mutual dependence of these events.

In accord with [6, 4.1], any one of the three pairs of events A_i, A_j, $1 \leq i < j \leq 3$, produces a joint experiment $\mathfrak{J}_{ij}$ whose probability distribution satisfies the linear system (3). Since the partition $\mathfrak{J}_3$ of the sample space is finer than each partition $\mathfrak{J}_{ij}$, its probability distribution $(\xi_1, \ldots, \xi_8)$ satisfies the linear system (5). After fixing the probabilities $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ of the components of *Yule's triple* $A = (A_1, A_2, A_3)$, the general solution of the last system depends on four parameters $\theta = (\theta_0, \ldots, \theta_3)$ chosen among ξ_k's. Taking into account that $\xi_k(\theta)$'s are probabilities, we obtain that θ varies within a subset $I_7(\alpha)$ of $\mathbb{R}^4$, which is described in Theorem 1. In case $\alpha \in (0,1)^3$ the set $I_7(\alpha)$ is a polytope, see [2, Ch. 12]. Since the system of linear inequalities (9) which define the polytope $I_7(\alpha)$ is minimal (Lemma 2), we can apply the machinery from the previous citation in order to use the corresponding properties of this polytope.

The 7-tuples (α, θ) vary within a polytope $I_7 \subset \mathbb{R}^7$ which is the inverse image of the 7-dimensional simplex Δ_7 via the affine isomorphism (7). The projection $p(\alpha, \theta) = \alpha$ produces the fibre bundle $(I_7, p, [0,1]^3)$ with fibre $p^{-1}(\alpha) = C_7(\alpha)$ where $C_7(\alpha) = \{\alpha\} \times I_7(\alpha)$, for the definition see [5, Part I, 2, 1.1]. This fibre bundle is used for classification of all equivalence classes of Yule's triples with given α and θ, cf. [6, Theorem 1]. An isomorphic fibre bundle can be used for classification of all probability distributions produced by the above equivalence classes of Yule's triples. The general patterns of these two fibre bundles are described in terms of very elementary algebraic geometry at the end of Subsection 4.2 where also classification Theorem 2 is formulated.

Corollary 1, (ii), yields that $0 < \xi_k(\theta) < 1$, $k = 1, \dots, 8$, if and only if $\theta \in \mathring{I}_7(\alpha)$. In particular, $\mathring{I}_7(\alpha)$ is the natural domain of the entropy function $E_\alpha(\theta)$ of the probability distribution $(\xi_k(\theta))_{k=1}^8$, defined in (11).

In Lemma 4 we prove that $E_\alpha(\theta)$ is a strictly concave function that can be extended in a unique way as continuous at the polytope $I_7(\alpha)$. Moreover, its continuous extension $\hat{E}_\alpha$ is also a strictly concave function. In Corollary 2 we show that all permutations of the members of Yule's triple $A = (A_1, A_2, A_3)$ have the same entropy.

Subsection 5.2 is devoted to finding the set of critical points of the entropy function $E_\alpha(\theta)$. It turns out that this set is not empty: The special point $\theta^{(\alpha)} \in \mathring{I}_7(\alpha)$ defined by the formulae (10) is critical, see Lemma 6.

Since the Hessian of $E_\alpha(\theta)$ is a negative definite quadratic form everywhere in its domain $\mathring{I}_7(\alpha)$, we obtain that the set of local maximums of the entropy function $E_\alpha(\theta)$ coincides with the set of its critical points, see Lemma 7.

In accord with Weierstrass theorem, the extended entropy function $\hat{E}_\alpha(\theta)$ attains an absolute maximum and an absolute minimum in its compact domain $I_7(\alpha)$. Theorems 3 and 4 make this statement more precise. The former asserts that $\hat{E}_\alpha(\theta)$ has a unique absolute maximum at the point $\theta^{(\alpha)}$. The latter uses the structure of the frontier of the polytope $I_7(\alpha)$, described, for example, in [2, Chapter 12, 12.1], and shows that $\hat{E}_\alpha(\theta)$ attains its absolute minimum only at some of its vertices. We note here an analogy with the simplex method.

Subsection 6.1 contains two statements that motivate the use the extended entropy function $\hat{E}_\alpha(\theta)$ for measuring the power of mutual relations among three events. In Lemma 8 we show that the components of a Yule's triple are mutually independent if and only if the corresponding θ coincides with $\theta^{(\alpha)}$. In other words, we observe mutual independence exactly when $\hat{E}_\alpha(\theta)$ attains its absolute maximum, which is in keeping conformity with our intuition. In the case of sample space with equally likely outcomes, Lemma 9 establishes the set-theoretic relations among the components of a Yule's triple when the corresponding θ lies on any one of the 3-faces of the polytope $I_7(\alpha)$. Intuitively, the "maximally" tight-fitting is observed at the vertices some of which are points of absolute minimum of $\hat{E}_\alpha(\theta)$.

Let $A = (A_1, A_2, A_3)$ be a Yule's triple with $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, $\alpha_1 = \Pr(A_1)$, $\alpha_2 = \Pr(A_2)$, $\alpha_3 = \Pr(A_3)$. In the final Subsection 6.2 we compose the extended entropy function $\hat{E}_\alpha(\theta)$ with a linear function and define a function $e_\alpha \colon I_7(\alpha) \to [0,1]$, whose value at any $\theta \in I_7(\alpha)$ corresponding to A is said to be degree of dependence of the events A_1, A_2, A_3. Note that $e_\alpha(\theta^{(\alpha)}) = 0$ (the events A_1, A_2, A_3 are mutually independent) and $e_\alpha(\theta_1) = 1$ for any vertex θ_1 where $\hat{E}_\alpha(\theta)$ attains its absolute minimum (the events A_1, A_2, A_3 are maximally dependent).

2. Definitions and Notation

Let $(\Omega, \mathcal{A}, \Pr)$ be a probability space with set of outcomes Ω, σ-algebra $\mathcal{A}$, and probability function Pr. In this paper we are using only the structure of Boolean algebra on $\mathcal{A}$.

We introduce the following notation:

Given events A_1, A_2, A_3 from $\mathcal{A}$, we set $A = (A_1, A_2, A_3) \in \mathcal{A}^3$;

R is the range of the probability function $\Pr\colon \mathcal{A} \to \mathbb{R}$;
Given $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$, we set $\alpha = (\alpha_1, \alpha_2, \alpha_3)$;
Given $\theta_0, \theta_1, \theta_2, \theta_3 \in \mathbb{R}$, we set $\theta = (\theta_0, \theta_1, \theta_2, \theta_3)$;
$I(\alpha_i, \alpha_j) = [\max(0, \alpha_i + \alpha_j - 1), \min(\alpha_i, \alpha_j)]$, $1 \leq i < j \leq 3$, see [6, 4.1];
$I^{(\alpha_i, \alpha_j)} = [\max(0, \alpha_i - \alpha_j), \min(\alpha_i, 1 - \alpha_j)]$, $1 \leq i < j \leq 3$;
$[(\alpha)]$ is the fiber of the surjective map

$$\mathcal{A}^3 \to R^3, (A_1, A_2, A_3) \mapsto (\Pr(A_1), \Pr(A_2), \Pr(A_3)),$$

over $\alpha \in R^3$;
$[(\alpha_i, \alpha_j)]$ is the fiber of the surjective map

$$\mathcal{A}^2 \to R^2, (A_i, A_j) \mapsto (\Pr(A_i), \Pr(A_j)),$$

over $(\alpha_i, \alpha_j) \in R^2$, $1 \leq i < j \leq 3$;

$$\theta_0^{(A)} = \Pr(A_1 \cap A_2 \cap A_3), \theta_1^{(A)} = \Pr(A_1^c \cap A_2 \cap A_3),$$

$$\theta_2^{(A)} = \Pr(A_1 \cap A_2^c \cap A_3), \theta_3^{(A)} = \Pr(A_1 \cap A_2 \cap A_3^c), A \in \mathcal{A}^3;$$

$$\theta^{(A)} = (\theta_0^{(A)}, \theta_1^{(A)}, \theta_2^{(A)}, \theta_3^{(A)});$$

$[(\alpha, \theta)]$ is the fiber of the map $[(\alpha)] \to R^4, A \mapsto \theta^{(A)}$, over any $\theta \in R^4$, and $R^{(\alpha)}$ is its range.

We note that the fibers $[(\alpha)]$ for $(\alpha) \in R^3$ form a partition of $\mathcal{A}^3$ and the fibers $[(\alpha, \theta)]$ for $\theta \in R^{(\alpha)}$ form a partition of $[(\alpha)]$.

The members of the fiber $[(\alpha)]$ are said to be *Yule's triples of type* (α). The members of the fiber $[(\alpha, \theta)]$ are called *Yule's triples of type* (α, θ).

3. Methods

In this paper we are using fundamentals of:
- Linear algebra,
- Affine geometry,
- Polytope theory,
- Fibre bundles,
- Real algebraic geometry.

4. Classification of Yule's Triples and Their Probability Distributions

4.1. The Probability Distribution of a Yule's Triple

Any ordered triple $A = (A_1, A_2, A_3) \in \mathcal{A}^3$ produces three experiments of the form

$$\mathfrak{J}_{ij} = (A_i \cap A_j) \cup (A_i \cap A_j^c) \cup (A_i^c \cap A_j) \cup (A_i^c \cap A_j^c), 1 \leq i < j \leq 3,$$

and the experiment

$$\mathfrak{J}_3 = (A_1 \cap A_2 \cap A_3) \cup (A_1^c \cap A_2 \cap A_3) \cup (A_1 \cap A_2^c \cap A_3) \cup (A_1 \cap A_2 \cap A_3^c) \cup$$

$$(A_1 \cap A_2^c \cap A_3^c) \cup (A_1^c \cap A_2 \cap A_3^c) \cup (A_1^c \cap A_2^c \cap A_3) \cup (A_1^c \cap A_2^c \cap A_3^c)$$

(cf. [8, I,§5]). We introduce the following notation:

$$\xi_1^{(A_i, A_j)} = \Pr(A_i \cap A_j), \xi_2^{(A_i, A_j)} = \Pr(A_i \cap A_j^c),$$

$$\xi_3^{(A_i,A_j)} = \Pr(A_i^c \cap A_j), \xi_4^{(A_i,A_j)} = \Pr(A_i^c \cap A_j^c), 1 \leq i < j \leq 3.$$

Moreover, we set

$$\begin{aligned}
\xi_1^{(A)} &= \Pr(A_1 \cap A_2^c \cap A_3^c), \xi_2^{(A)} = \Pr(A_1^c \cap A_2 \cap A_3^c),\\
\xi_3^{(A)} &= \Pr(A_1^c \cap A_2^c \cap A_3), \xi_4^{(A)} = \Pr(A_1^c \cap A_2^c \cap A_3^c),\\
\xi_5^{(A)} &= \Pr(A_1 \cap A_2 \cap A_3), \xi_6^{(A)} = \Pr(A_1^c \cap A_2 \cap A_3),\\
\xi_7^{(A)} &= \Pr(A_1 \cap A_2^c \cap A_3), \xi_8^{(A)} = \Pr(A_1 \cap A_2 \cap A_3^c).
\end{aligned} \tag{1}$$

The above probabilities satisfy the following identities:

$$\begin{aligned}
&\xi_5^{(A)} + \xi_8^{(A)} = \xi_1^{(A_1,A_2)}, \xi_1^{(A)} + \xi_7^{(A)} = \xi_2^{(A_1,A_2)},\\
&\xi_2^{(A)} + \xi_6^{(A)} = \xi_3^{(A_1,A_2)}, \xi_3^{(A)} + \xi_4^{(A)} = \xi_4^{(A_1,A_2)},\\
&\xi_5^{(A)} + \xi_7^{(A)} = \xi_1^{(A_1,A_3)}, \xi_1^{(A)} + \xi_8^{(A)} = \xi_2^{(A_1,A_3)},\\
&\xi_3^{(A)} + \xi_6^{(A)} = \xi_3^{(A_1,A_3)}, \xi_2^{(A)} + \xi_4^{(A)} = \xi_4^{(A_1,A_3)},\\
&\xi_5^{(A)} + \xi_6^{(A)} = \xi_1^{(A_2,A_3)}, \xi_2^{(A)} + \xi_8^{(A)} = \xi_2^{(A_2,A_3)},\\
&\xi_3^{(A)} + \xi_7^{(A)} = \xi_3^{(A_2,A_3)}, \xi_1^{(A)} + \xi_4^{(A)} = \xi_4^{(A_2,A_3)}.
\end{aligned} \tag{2}$$

For any $1 \leq i < j \leq 3$ and any $(A_i, A_j) \in [(\alpha_i, \alpha_j)]$, the probability distribution

$$(\xi_1^{(i,j)}, \xi_2^{(i,j)}, \xi_3^{(i,j)}, \xi_4^{(i,j)}) = (\xi_1^{(A_i,A_j)}, \xi_2^{(A_i,A_j)}, \xi_3^{(A_i,A_j)}, \xi_4^{(A_i,A_j)})$$

satisfies the linear system

$$\left|\begin{array}{ccccccccc}
\xi_1^{(i,j)} & + & \xi_2^{(i,j)} & & & & & = & \alpha_i\\
 & & & & \xi_3^{(i,j)} & + & \xi_4^{(i,j)} & = & 1-\alpha_i\\
\xi_1^{(i,j)} & & & + & \xi_3^{(i,j)} & & & = & \alpha_j\\
 & & \xi_2^{(i,j)} & & & + & \xi_4^{(i,j)} & = & 1-\alpha_j.
\end{array}\right. \tag{3}$$

The identities (2) and the linear systems (3) yield that for any ordered triple $A \in [\alpha]$, the probability distribution

$$(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7, \xi_8) = (\xi_1^{(A)}, \xi_2^{(A)}, \xi_3^{(A)}, \xi_4^{(A)}, \xi_5^{(A)}, \xi_6^{(A)}, \xi_7^{(A)}, \xi_8^{(A)}) \tag{4}$$

satisfies the linear system

$$\left|\begin{array}{ccccccccccccccccc}
\xi_1 & & & & & & & + & \xi_5 & & & + & \xi_7 & + & \xi_8 & = & \alpha_1\\
 & & \xi_2 & + & \xi_3 & + & \xi_4 & & & + & \xi_6 & & & & & = & 1-\alpha_1\\
 & & \xi_2 & & & & & + & \xi_5 & + & \xi_6 & & & + & \xi_8 & = & \alpha_2\\
\xi_1 & & & + & \xi_3 & + & \xi_4 & & & & & + & \xi_7 & & & = & 1-\alpha_2\\
 & & & & \xi_3 & & & + & \xi_5 & + & \xi_6 & + & \xi_7 & & & = & \alpha_3\\
\xi_1 & + & \xi_2 & & & + & \xi_4 & & & & & & & + & \xi_8 & = & 1-\alpha_3.
\end{array}\right. \tag{5}$$

Let us denote for short $\xi = (\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7, \xi_8)$ and let H_7 be the affine hyperplane in $\mathbb{R}^8$ with equation $\xi_1 + \xi_2 + \xi_3 + \xi_4 + \xi_5 + \xi_6 + \xi_7 + \xi_8 = 1$. For any $\alpha \in \mathbb{R}^3$ the solutions of (5) depend on four parameters, say $\theta_0 = \xi_5$, $\theta_1 = \xi_6$, $\theta_2 = \xi_7$, $\theta_3 = \xi_8$, and for any triple $\alpha \in \mathbb{R}^3$ form a 4-dimensional affine space ℓ_α in H_7 with parametric representation

$$\ell_\alpha\colon \left\{\begin{array}{lcccccccccccccccc} \xi_1 & = & & & \alpha_1 & & & & & - & \theta_0 & & & - & \theta_2 & - & \theta_3 \\ \xi_2 & = & & & & & \alpha_2 & & & - & \theta_0 & - & \theta_1 & & & - & \theta_3 \\ \xi_3 & = & & & & & & & \alpha_3 & - & \theta_0 & - & \theta_1 & - & \theta_2 & & \\ \xi_4 & = & 1 & - & \alpha_1 & - & \alpha_2 & - & \alpha_3 & + & 2\theta_0 & + & \theta_1 & + & \theta_2 & + & \theta_3 \\ \xi_5 & = & & & & & & & & & \theta_0 & & & & & & \\ \xi_6 & = & & & & & & & & & & & \theta_1 & & & & \\ \xi_7 & = & & & & & & & & & & & & & \theta_2 & & \\ \xi_8 & = & & & & & & & & & & & & & & & \theta_3 \end{array}\right. \tag{6}$$

The map

$$\iota_7\colon \mathbb{R}^7 \to H_7, (\alpha, \theta) \mapsto \xi, \tag{7}$$

defined by formulae (6) is an affine isomorphism with inverse affine isomorphism

$$\chi_7\colon H_7 \to \mathbb{R}^7, \xi \mapsto (\xi_1 + \xi_5 + \xi_7 + \xi_8, \xi_2 + \xi_5 + \xi_6 + \xi_8, \xi_3 + \xi_5 + \xi_6 + \xi_7, \xi_5, \xi_6, \xi_7, \xi_8). \tag{8}$$

The symmetric group S_3 acts on $\mathbb{R}^7$ by the rule $\sigma(\alpha, \theta) = (\sigma\alpha; \sigma\theta)$, where $\sigma\alpha = (\alpha_{\sigma^{-1}(1)}, \alpha_{\sigma^{-1}(2)}, \alpha_{\sigma^{-1}(3)})$ and $\sigma\theta = (\theta_0, \theta_{\sigma^{-1}(1)}, \theta_{\sigma^{-1}(2)}, \theta_{\sigma^{-1}(3)})$, $\sigma \in S_3$. When necessary, we write σ_α and σ_θ in order to distinguish the actions of σ on α's and θ's, respectively.

On the other hand, we transport the action of S_3 on the set $\{6, 7, 8\}$ via the bijection $1 \mapsto 6, 2 \mapsto 7, 3 \mapsto 8$ and define an action of S_3 on the hyperplane H_7 by the formula

$$\sigma\xi = (\xi_{\sigma^{-1}(1)}, \xi_{\sigma^{-1}(2)}, \xi_{\sigma^{-1}(3)}, \xi_4, \xi_5, \xi_{\sigma^{-1}(6)}, \xi_{\sigma^{-1}(7)}, \xi_{\sigma^{-1}(8)}).$$

Lemma 1. The affine isomorphism ι_7 is also an isomorphism of S_3-sets: $\iota_7(\sigma(\alpha, \theta)) = \sigma\iota_7(\alpha, \theta)$.

Proof. We check the statement for a set of generators of S_3: For $\sigma = (12)$ we have

$$\xi_1((12)(\alpha, \theta)) = \xi_2(\alpha, \theta), \xi_2((12)(\alpha, \theta)) = \xi_1(\alpha, \theta),$$

$$\xi_6((12)(\alpha, \theta)) = \xi_7(\alpha, \theta), \xi_7((12)(\alpha, \theta)) = \xi_6(\alpha, \theta).$$

For $\sigma = (23)$ we have

$$\xi_2((23)(\alpha, \theta)) = \xi_3(\alpha, \theta), \xi_3((23)(\alpha, \theta)) = \xi_2(\alpha, \theta),$$

$$\xi_7((23)(\alpha, \theta)) = \xi_8(\alpha, \theta), \xi_8((23)(\alpha, \theta)) = \xi_7(\alpha, \theta).$$

■

4.2. The Geometric Classification

After fixing the coordinates α_1, α_2, and α_3, the isomorphism ι_7 from (7) maps the 4-dimensional affine space $\zeta_\alpha = \{\alpha\} \times \mathbb{R}^4$ onto the 4-dimensional affine space ℓ_α in H_7. We denote by $\iota_7^{(\alpha)}$ the (affine) restriction of ι_7 on ζ_α, so $\iota_7^{(\alpha)}\colon \zeta_\alpha \to \ell_\alpha$.

The trace of the 8-dimensional cube $\{\xi \in \mathbb{R}^8 | 0 \leq \xi_k \leq 1, k = 1, \dots, 8\}$ onto the hyperplane H_7 is the 7-dimensional simplex Δ_7 defined in H_7 by the inequalities $\xi_1 \geq 0, \dots, \xi_8 \geq 0$. The inverse image $T_7 = \iota_7^{-1}(\Delta_7)$ via the affine isomorphism ι_7 is the convex polyhedron in $\mathbb{R}^7$ with non-empty interior, defined by the system of inequalities

$$T_7\colon \left\{\begin{array}{lcccccccl} \theta_0 & & & + & \theta_2 & + & \theta_3 & \leq & \alpha_1 \\ \theta_0 & + & \theta_1 & & & + & \theta_3 & \leq & \alpha_2 \\ \theta_0 & + & \theta_1 & + & \theta_2 & & & \leq & \alpha_3 \\ 2\theta_0 & + & \theta_1 & + & \theta_2 & + & \theta_3 & \geq & \alpha_1 + \alpha_2 + \alpha_3 - 1. \\ \theta_0 & & & & & & & \geq & 0 \\ & & \theta_1 & & & & & \geq & 0 \\ & & & & \theta_2 & & & \geq & 0 \\ & & & & & & \theta_3 & \geq & 0 \end{array}\right. \tag{9}$$

The form (8) of the inverse isomorphism χ_7 yields that $T_7 \subset [0,1]^7$. In particular, T_7 is a polytope. Note that we are using the terminology about polytopes introduced in [2, Ch. 12].

For any $\alpha \in \mathbb{R}^3$ we set $C_7(\alpha) = \zeta_\alpha \cap T_7$, so $C_7(\alpha) = \{\alpha\} \times I_7(\alpha)$, where $I_7(\alpha) \subset \mathbb{R}^4$ and $\mathbb{R}^4$ is furnished with coordinates θ. The subset $I_7(\alpha)$ is defined in $\mathbb{R}^4$ via the system (9) with fixed α. Hence $I_7(\alpha)$ is a convex bounded polyhedron in $\mathbb{R}^4$. We also set $D_7(\alpha) = \iota_7(C_7(\alpha))$. Since $\iota_7(\zeta_\alpha) = \ell_\alpha$, we obtain that $D_7(\alpha) = \ell_\alpha \cap \Delta_7$.

We consider T_7, $\zeta_\alpha \simeq \mathbb{R}^4$, $C_7(\alpha)$, $I_7(\alpha)$, ℓ_α, Δ_7, and $D_7(\alpha)$ as topological subspaces of the corresponding ambient linear spaces, with topology induced by their standard topology. Moreover, for each subset A of a topological space X we denote by $\mathring{A}$ its interior with respect to X. We note that $\mathring{A}$ is the largest open set contained in A, see [3, § 1, $n^o 6$].

Lemma 2. The minimal number of half-spaces in $\mathbb{R}^4$, whose intersection is the polyhedron $I_7(\alpha)$ is 8.

Proof. We can not omit any one of the inequalities in (9) formed by the free variables $\xi_5 = \theta_0$, $\xi_6 = \theta_1$, $\xi_7 = \theta_2$, and $\xi_8 = \theta_3$. It turns out that the general solution of the linear system (5) can also be written in terms of the free variables ξ_1, ξ_2, ξ_3, and ξ_4. In particular, neither of the inequalities $\xi_1 \geq 0$, $\xi_2 \geq 0$, $\xi_3 \geq 0$, and $\xi_4 \geq 0$, that define the polytope T_7 can be omitted, too. ■

We define the point $\theta^{(\alpha)} \in \mathbb{R}^4$ by the formulae

$$\theta_0^{(\alpha)} = \alpha_1\alpha_2\alpha_3, \theta_1^{(\alpha)} = (1-\alpha_1)\alpha_2\alpha_3, \theta_2^{(\alpha)} = \alpha_1(1-\alpha_2)\alpha_3, \theta_3^{(\alpha)} = \alpha_1\alpha_2(1-\alpha_3). \tag{10}$$

Lemma 3. If $\alpha \in [0,1]^3$, then $\theta^{(\alpha)} \in I_7(\alpha)$ and the following three statements are equivalent:
(i) One has $\alpha \in (0,1)^3$.
(ii) One has $\theta^{(\alpha)} \in \mathring{I}_7(\alpha)$.
(iii) One has $\mathring{I}_7(\alpha) \neq \emptyset$.

Proof. The equalities $\theta_1 + \theta_3 + \theta_4 - \alpha_1 = -\alpha_1(1-\alpha_2)(1-\alpha_3)$, $\theta_1 + \theta_2 + \theta_4 - \alpha_2 = -\alpha_2(1-\alpha_1)(1-\alpha_3)$, $\theta_1 + \theta_2 + \theta_3 - \alpha_3 = -\alpha_3(1-\alpha_1)(1-\alpha_2)$, and $2\theta_1 + \theta_2 + \theta_3 + \theta_4 - \alpha_1 - \alpha_2 - \alpha_3 + 1 = (1-\alpha_1)(1-\alpha_2)(1-\alpha_3)$ yield that the system (9) is satisfied if $\alpha \in [0,1]^3$. If, in addition, $\alpha \in (0,1)^3$, then (9) with strict inequalities holds. Thus, the implication (i) $\Longrightarrow$ (ii) is also proved.

(ii) $\Longrightarrow$ (iii) This is trivial.

(iii) $\Longrightarrow$ (i) Let $\theta \in \mathring{I}_7(\alpha)$. Then $\xi_k(\theta) > 0$, $k = 1, \ldots, 8$, their sum is 1, and satisfy the linear system (5). Therefore $\alpha \in (0,1)^3$. ■

Theorem 1. (i) One has

$$I_7(\alpha) = \begin{cases} (0,0,0,0) & \text{if at least two of } \alpha_i's \text{ are } 0 \\ \{0\} \times I(\alpha_2,\alpha_3) \times \{0\} \times \{0\} & \text{if } \alpha_1 = 0, \alpha_2 > 0, \alpha_3 > 0 \\ \{0\} \times \{0\} \times I(\alpha_1,\alpha_3) \times \{0\} & \text{if } \alpha_2 = 0, \alpha_1 > 0, \alpha_3 > 0 \\ \{0\} \times \{0\} \times \{0\} \times I(\alpha_1,\alpha_2) & \text{if } \alpha_3 = 0, \alpha_1 > 0, \alpha_2 > 0 \\ \{\alpha_3\} \times \{0\} \times \{0\} \times \{1-\alpha_3\} & \text{if } \alpha_1 = 1, \alpha_2 = 1, \alpha_3 > 0 \\ \{\alpha_2\} \times \{0\} \times \{1-\alpha_2\} \times \{0\} & \text{if } \alpha_1 = 1, \alpha_3 = 1, \alpha_2 > 0 \\ \{\alpha_1\} \times \{1-\alpha_1\} \times \{0\} \times \{0\} & \text{if } \alpha_2 = 1, \alpha_3 = 1, \alpha_1 > 0 \\ \{(\alpha_2-\theta_3, 0, \alpha_3-\alpha_2+\theta_3, \theta_3) | \theta_3 \in I^{(\alpha_2,\alpha_3)}\} & \text{if } \alpha_1 = 1, \alpha_2 > 0, \alpha_3 > 0 \\ \{(\alpha_3-\theta_1, \theta_1, 0, \alpha_1-\alpha_3+\theta_1) | \theta_1 \in I^{(\alpha_3,\alpha_1)}\} & \text{if } \alpha_2 = 1, \alpha_1 > 0, \alpha_3 > 0 \\ \{(\alpha_1-\theta_2, \alpha_2-\alpha_1+\theta_2, \theta_2, 0) | \theta_2 \in I^{(\alpha_1,\alpha_2)}\} & \text{if } \alpha_3 = 1, \alpha_1 > 0, \alpha_2 > 0 \end{cases}$$

and $I_7(\alpha)$ is a polytope in $\mathbb{R}^4$ if $\alpha \in (0,1)^3$.

(ii) One has $\iota_7(\mathring{C}_7(\alpha)) = \mathring{D}_7(\alpha)$ the interiors being with respect to affine spaces ζ_α and ℓ_α, respectively.

Proof. (i) The systems (5) and (9) imply the equalities. In case $\alpha \in (0,1)^3$, Lemma 3 yields that the bounded convex polyhedron $I_7(\alpha)$ in $\mathbb{R}^4$ has non-empty interior. In other words, it is a polytope.

(ii) It is enough to note that the (affine) restriction $\iota_7^{(\alpha)}\colon \zeta_\alpha \to \ell_\alpha$ is, in particular, a homeomorphism.

∎

Corollary 1. Let $\alpha \in \mathbb{R}^3$.

(i) The system of constraint conditions $0 \leq \xi_k(\theta) \leq 1, k = 1, \ldots, 8$, on the solutions (6) of linear system (5) is equivalent to the property $\theta \in I_7(\alpha)$.

(ii) One has $0 < \xi_k(\theta) < 1$, $k = 1, \ldots, 8$, if and only if $\theta \in \mathring{I}_7(\alpha)$.

Proof. (i) The equalities $C_7(\alpha) = \zeta_\alpha \cap T_7$ and $D_7(\alpha) = \ell_\alpha \cap \Delta_7$ imply part (i). We have $\mathring{C}_7(\alpha) = \zeta_\alpha \cap \mathring{T}_7$ and $\mathring{D}_7(\alpha) = \ell_\alpha \cap \mathring{\Delta}_7$, where the interiors $\mathring{T}_7$ and $\mathring{\Delta}_7$ are with respect to affine spaces $\mathbb{R}^7$ and H_7, respectively. Now, Theorem 1, (ii), yields part (ii).

∎

We have $R^{(\alpha)} \subset I_7(\alpha)$ and define $I_7^{(\cdot)}(\alpha) = R^{(\alpha)}$. The *dotted polytope* $C_7^{(\cdot)}(\alpha) = \{\alpha\} \times I_7^{(\cdot)}(\alpha)$, $(\alpha) \in R^3$, is the locus of all 7-tuples of probabilities $(\alpha, \theta^{(A)})$, where $A \in [(\alpha)]$.

By plugging $\theta^{(\alpha)}$ in the formulae (6), we obtain the point $\xi^{(\alpha)} \in H_7$ with coordinates

$$\xi_1^{(\alpha)} = \alpha_1(1-\alpha_2)(1-\alpha_3), \xi_2^{(\alpha)} = (1-\alpha_1)\alpha_2(1-\alpha_3),$$

$$\xi_3^{(\alpha)} = (1-\alpha_1)(1-\alpha_2)\alpha_3, \xi_4^{(\alpha)} = (1-\alpha_1)(1-\alpha_2)(1-\alpha_3),$$

$$\xi_5^{(\alpha)} = \alpha_1\alpha_2\alpha_3, \xi_6^{(\alpha)} = (1-\alpha_1)\alpha_2\alpha_3, \xi_7^{(\alpha)} = \alpha_1(1-\alpha_2)\alpha_3, \xi_8^{(\alpha)} = \alpha_1\alpha_2(1-\alpha_3).$$

Let U_3 be the rational 3-dimensional algebraic manifold defined in $\mathbb{R}^7$ by the equations (10). In other words, U_3 is the locus of the points in $\mathbb{R}^7$ of the form $(\alpha, \theta^{(\alpha)})$, $\alpha \in \mathbb{R}^3$. Let us denote $W_3 = \iota_7(U_3)$, so W_3 is the locus of the points $\xi^{(\alpha)}$, $\alpha \in \mathbb{R}^3$, in H_7. Then $\chi_7(W_3) = U_3$, W_3 is an algebraic subvariety of H_7, and the restrictions of ι_7 and χ_7 on U_3 and W_3, respectively, form a pair of mutually inverse isomorphisms of 3-dimensional rational algebraic manifolds. Moreover, $W_3 \cap \ell_\alpha = \{\xi^{(\alpha)}\}$ for any $\alpha \in \mathbb{R}^3$. Let us denote $\kappa_3 = \iota_3 \circ \delta_3$, where δ_3 is the isomorphism of algebraic manifolds $\mathbb{R}^3 \to U_3$, $\alpha \mapsto (\alpha, \theta^{(\alpha)})$. Therefore, $\kappa_3\colon \mathbb{R}^3 \to W_3$ is also an isomorphism of algebraic manifolds.

We have the product vector bundle with total space $\mathbb{R}^7$, base $\mathbb{R}^3$, projection $(\alpha, \theta) \mapsto \alpha$, and fibre ζ_α. Now, we transport the structure of fibre bundle by means of the pair of isomorphisms (ι_7, κ_3) to H_7 and W_3, thus obtaining a structure of vector bundle with total space H_7, base W_3, projection $\pi\colon H_7 \to W_3$, with $\pi^{-1}(\xi^{(\alpha)}) = \ell_\alpha$. Via restriction we obtain a fibre bundle with total space T_7, base $[0,1]^3$, projection $(\alpha, \theta) \mapsto \alpha$, and fibre $C_7(\alpha)$, as well as a fibre bundle with total space Δ_7 and base $w_3 = \kappa_3([0,1]^3)$. Combining the equality $\iota_7(C_7(\alpha)) = D_7(\alpha)$, Lemma 3, and Theorem 1, (ii), we obtain that if $\alpha \in [0,1]^3$ (respectively, $\alpha \in (0,1)^3$), then $\xi^{(\alpha)} \in D_7(\alpha)$ (respectively, $\xi^{(\alpha)} \in \mathring{D}_7(\alpha)$). Thus, $w_3 \cap D_7(\alpha) = \{\xi^{(\alpha)}\}$ and the projection $\pi\colon \Delta_7 \to w_3$ has fibres $\pi^{-1}(\xi^{(\alpha)}) = D_7(\alpha)$. Moreover, the restriction of the pair (ι_7, κ_3) is an isomorphism of fibre bundles.

For the sake of transparency, we note that $T_7 = \cup_{(\alpha)\in[0,1]^3} C_7(\alpha)$, $\Delta_7 = \cup_{(\alpha)\in[0,1]^3} D_7(\alpha)$. The unions $T_7^{(\cdot)} = \cup_{(\alpha)\in R^3} C_7^{(\cdot)}(\alpha)$, $\Delta_7^{(\cdot)} = \cup_{(\alpha)\in R^3} D_7^{(\cdot)}(\alpha)$ are the corresponding *dotted polytopes*.

The above considerations yield the following classification theorem:

Theorem 2. (i) The affine isomorphism $\iota_7\colon \mathbb{R}^7 \to H_7$ transforms any polytope $C_7(\alpha)$ (resp., dotted polytope $C_7^{(\cdot)}(\alpha)$) onto the polytope $D_7(\alpha)$ (resp., onto the dotted polytope $D_7^{(\cdot)}(\alpha)$).

(ii) The dotted polytope $C_7^{(\cdot)}(\alpha)$ is the classification space of all Yule's triples of type $[(\alpha, \theta)]$. The dotted polytope $\Delta_7^{(\cdot)}(\alpha)$ is the classification space of all probability distributions (1) produced by Yule's triples of type $[(\alpha, \theta)]$.

(iii) ι_7 maps the polytope T_7 (resp., dotted polytope $T_7^{(\cdot)}$) onto the polytope Δ_7 (resp., onto the dotted polytope $\Delta_7^{(\cdot)}$).

(iv) The dotted polytope $T_7^{(\cdot)}$ is the classification space of all Yule's triples. The dotted polytope $\Delta_7^{(\cdot)}$ is the classification space of all probability distributions produced by Yule's triples.

5. Entropy and Dependence of Yule's Triples

In this section we suppose $\alpha \in (0,1)^3$, that is (Lemma 3), $\mathring{I}_7(\alpha) \neq \emptyset$.

5.1. The Entropy Function

The function $E\colon \mathring{\Delta}_7 \to \mathbb{R}$, $E(\xi) = -\sum_{k=1}^{8} \xi_k \ln \xi_k$, is strictly concave since the open simplex $\mathring{\Delta}_7$ is convex and all of its "entropy" summands $E^{(k)}(\xi) = -\xi_k \ln \xi_k$ are strictly concave. Let us fix $\alpha \in (0,1)^3$ and let

$$E_\alpha(\theta) = \sum_{k=1}^{8} E_\alpha^{(k)}(\theta), E_\alpha^{(k)}(\theta) = -\xi_k(\theta) \ln \xi_k(\theta), \tag{11}$$

be the composition of E with the affine isomorphism $\iota_7^{(\alpha)}$: $E_\alpha(\theta) = E(\iota_7^{(\alpha)}(\theta))$. In accord with Corollary 1, (ii), the entropy function (11) of the experiment $\mathfrak{J}_3$ has $\mathring{I}_7(\alpha)$ as a natural domain: $E_\alpha\colon \mathring{I}_7(\alpha) \to \mathbb{R}$.

Lemma 4. (i) The entropy function E_α is a strictly concave function.

(ii) The entropy function E_α can be extended as continuous at $I_7(\alpha)$ and this extension $\hat{E}_\alpha$ is unique.

(iii) The continuous extension $\hat{E}_\alpha$ of E_α at $I_7(\alpha)$ is also a strictly concave function.

Proof. Note that the polytope $I_7(\alpha)$ and its interior $\mathring{I}_7(\alpha)$ are bounded convex sets.

(i) The function E_α is composition of the affine map $\iota_7^{(\alpha)}$ followed by the strictly concave function $E(\xi)$.

(ii) We apply [3, § 8, n^o5, Theorem 1].

(iii) The point $\theta^{(0)}$ belongs to the frontier of the polytope $I_7(\alpha)$ if and only if $\xi_k(\theta^{(0)}) = 0$ for indices k from some set K and $\xi_k(\theta^{(0)}) > 0$ for the rest of the indices, where $k = 1, \ldots, 8$. Moreover, for any $k \in K$ we have $E^{(k)}(\theta) \to 0$ when $\theta \to \theta^{(0)}$, $\theta \in \mathring{I}_7(\alpha)$. In other words, $\hat{E}^{(k)}(\theta^{(0)}) = 0$.

A boundary transition yields that $\hat{E}_\alpha$ is a concave function. Moreover, since there are indices $k \notin K$, the function E_α is strictly concave. Indeed, let $\theta^{(1)} \in \mathring{I}_7(\alpha)$ and $\lambda \in (0,1)$. In accord with [2, Ch. 11, Lemma 11.2.4], we have $(1-\lambda)\theta^{(0)} + \lambda\theta^{(1)} \in \mathring{I}_7(\alpha)$, hence

$$\hat{E}^{(k)}((1-\lambda)\theta^{(0)} + \lambda\theta^{(1)}) = E^{(k)}((1-\lambda)\theta^{(0)} + \lambda\theta^{(1)}) < (1-\lambda)E^{(k)}(\theta^{(0)}) + \lambda E^{(k)}(\theta^{(1)})$$

for any $k = 1, \ldots, 8$.

In case $k \notin K$ we have $\hat{E}^{(k)}(\theta^{(0)}) = E^{(k)}(\theta^{(0)})$ and we are done. Now, let $k \in K$ and let $\theta \to \theta^{(0)}$, $\theta \in \mathring{I}_7(\alpha)$. We obtain

$$\hat{E}^{(k)}((1-\lambda)\theta^{(0)} + \lambda\theta^{(1)}) = \lim_{\theta\to\theta^{(0)}} E^{(k)}((1-\lambda)\theta + \lambda\theta^{(1)}) \leq$$

$$(1-\lambda) \lim_{\theta\to\theta^{(0)}} E^{(k)}(\theta) + \lambda E^{(k)}(\theta^{(1)}) = (1-\lambda)\hat{E}^{(k)}(\theta^{(0)}) + \lambda\hat{E}^{(k)}(\theta^{(1)}).$$

■

The symmetric group S_3 acts on the entropy functions $E_\alpha(\theta)$ by the rule $\sigma E_\alpha(\theta) = E_\alpha(\sigma^{-1}\theta)$, $\sigma \in S_3$.

Lemma 5. If $\sigma \in S_3$, then $E_{\sigma\alpha}(\theta) = \sigma E_\alpha(\theta)$ and $I_7(\sigma\alpha) = \sigma_\theta I_7(\alpha)$.

Proof. (i) According to Lemma 1, we have $\sigma^{-1}E_{\sigma\alpha}(\theta) = E_{\sigma\alpha}(\sigma\theta) = E(\iota_7^{(\sigma\alpha)}(\sigma\theta)) = E(\iota_7(\sigma\alpha, \sigma\theta)) = E(\sigma\iota_7(\alpha, \theta)) = E(\iota_7(\alpha, \theta)) = E_\alpha(\theta)$. Finally, the domain of $\sigma E_\alpha(\theta)$ is the polytope $\sigma_\theta I_7(\alpha)$ and we obtain $I_7(\sigma\alpha) = \sigma_\theta I_7(\alpha)$.

■

Corollary 2. Let $\sigma \in S_3$.

(i) One has $\hat{E}_{\sigma\alpha}(\sigma\theta) = \hat{E}_\alpha(\theta)$.

(ii) All permutations of the members of Yule's triple $A = (A_1, A_2, A_3)$ have the same entropy: If $A \in [(\alpha)]$, then $\sigma A \in [(\sigma\alpha)]$ and $\hat{E}_{\sigma\alpha}(\theta^{(\sigma A)}) = \hat{E}_\alpha(\theta^{(A)})$.

Proof. (i) Let $\theta^{(0)}$ be point from the frontier of the polytope $I_7(\alpha)$. Then $\sigma\theta^{(0)}$ is point from the frontier of the polytope $I_7(\sigma\alpha)$ with interior $\sigma_\theta \mathring{I}_7(\alpha)$. We have $\theta \to \theta^{(0)}$, $\theta \in \mathring{I}_7(\alpha)$, if and only if $\sigma\theta \to \sigma\theta^{(0)}$, $\sigma\theta \in \sigma\mathring{I}_7(\alpha)$. The equality from Lemma 5 can be written in the form $E_{\sigma\alpha}(\sigma\theta) = E_\alpha(\theta)$ and a boundary transition yields the result.

(ii) Implied by part (i).

■

5.2. The Entropy Function and its Critical Points

For any $\theta \in \mathring{I}_7(\alpha)$ we obtain

$$\frac{\partial E_\alpha(\theta)}{\partial\theta_0} = \ln\frac{\xi_1(\theta)\xi_2(\theta)\xi_3(\theta)}{\xi_4^2(\theta)\xi_5(\theta)}, \frac{\partial E_\alpha(\theta)}{\partial\theta_1} = \ln\frac{\xi_2(\theta)\xi_3(\theta)}{\xi_4(\theta)\xi_6(\theta)},$$

$$\frac{\partial E_\alpha(\theta)}{\partial\theta_2} = \ln\frac{\xi_1(\theta)\xi_3(\theta)}{\xi_4(\theta)\xi_7(\theta)}, \frac{\partial E_\alpha(\theta)}{\partial\theta_3} = \ln\frac{\xi_1(\theta)\xi_2(\theta)}{\xi_4(\theta)\xi_8(\theta)}.$$

Thus, the set of critical points of the function $E_\alpha(\theta)$ is the intersection of the interior $\mathring{I}_7(\alpha) \subset \mathbb{R}^4$ and the algebraic variety in $\mathbb{R}^4$ with equations

$$\xi_1(\theta)\xi_2(\theta)\xi_3(\theta) - \xi_4^2(\theta)\xi_5(\theta) = 0, \xi_2(\theta)\xi_3(\theta) - \xi_4(\theta)\xi_6(\theta) = 0,$$

$$\xi_1(\theta)\xi_3(\theta) - \xi_4(\theta)\xi_7(\theta) = 0, \xi_1(\theta)\xi_2(\theta) - \xi_4(\theta)\xi_8(\theta) = 0.$$

Lemma 6. (i) The point $\theta^{(\alpha)}$ is a critical point of the entropy function E_α.

(ii) One has

$$E_\alpha(\theta^{(\alpha)}) = -\ln\left(\alpha_1^{\alpha_1}\alpha_2^{\alpha_2}\alpha_3^{\alpha_3}(1-\alpha_1)^{1-\alpha_1}(1-\alpha_2)^{1-\alpha_2}(1-\alpha_3)^{1-\alpha_3}\right).$$

Proof. (i) We have

$$\xi_1^{(\alpha)}\xi_2^{(\alpha)}\xi_3^{(\alpha)} - \left(\xi_4^{(\alpha)}\right)^2\xi_5^{(\alpha)} =$$
$$\alpha_1(1-\alpha_2)(1-\alpha_3)(1-\alpha_1)\alpha_2(1-\alpha_3)(1-\alpha_1)(1-\alpha_2)\alpha_3 -$$
$$(1-\alpha_1)^2(1-\alpha_2)^2(1-\alpha_3)^2\alpha_1\alpha_2\alpha_3 = 0,$$
$$\xi_2^{(\alpha)}\xi_3^{(\alpha)} - \xi_4^{(\alpha)}\xi_6^{(\alpha)} =$$
$$(1-\alpha_1)\alpha_2(1-\alpha_3)(1-\alpha_1)(1-\alpha_2)\alpha_3 - (1-\alpha_1)(1-\alpha_2)(1-\alpha_3)(1-\alpha_1)\alpha_2\alpha_3 = 0,$$
$$\xi_1^{(\alpha)}\xi_3^{(\alpha)} - \xi_4^{(\alpha)}\xi_7^{(\alpha)} =$$
$$\alpha_1(1-\alpha_2)(1-\alpha_3)(1-\alpha_1)(1-\alpha_2)\alpha_3 - (1-\alpha_1)(1-\alpha_2)(1-\alpha_3)\alpha_1(1-\alpha_2)\alpha_3 = 0,$$
$$\xi_1^{(\alpha)}\xi_2^{(\alpha)} - \xi_4^{(\alpha)}\xi_8^{(\alpha)} =$$
$$\alpha_1(1-\alpha_2)(1-\alpha_3)(1-\alpha_1)\alpha_2(1-\alpha_3) - (1-\alpha_1)(1-\alpha_2)(1-\alpha_3)\alpha_1\alpha_2(1-\alpha_3) = 0.$$

(ii) We have

$$-E_\alpha(\theta^{(\alpha)}) = -E(\xi^{(\alpha)}) = \sum_{k=1}^{8}\xi_k^{(\alpha)}\ln\xi_k^{(\alpha)} =$$
$$\xi_1^{(\alpha)}\ln(\alpha_1(1-\alpha_2)(1-\alpha_3)) + \xi_2^{(\alpha)}\ln((1-\alpha_1)\alpha_2(1-\alpha_3)) +$$
$$\xi_3^{(\alpha)}\ln((1-\alpha_1)(1-\alpha_2)\alpha_3) + \xi_4^{(\alpha)}\ln((1-\alpha_1)(1-\alpha_2)(1-\alpha_3)) +$$
$$\xi_5^{(\alpha)}\ln(\alpha_1\alpha_2\alpha_3) + \xi_6^{(\alpha)}\ln((1-\alpha_1)\alpha_2\alpha_3) +$$
$$\xi_7^{(\alpha)}\ln(\alpha_1(1-\alpha_2)\alpha_3) + \xi_8^{(\alpha)}\ln(\alpha_1\alpha_2(1-\alpha_3)) =$$
$$(\xi_1^{(\alpha)} + \xi_5^{(\alpha)} + \xi_7^{(\alpha)} + \xi_8^{(\alpha)})\ln\alpha_1 + (\xi_2^{(\alpha)} + \xi_5^{(\alpha)} + \xi_6^{(\alpha)} + \xi_8^{(\alpha)})\ln\alpha_2 +$$

$$(\xi_3^{(\alpha)} + \xi_5^{(\alpha)} + \xi_6^{(\alpha)} + \xi_7^{(\alpha)}) \ln \alpha_3 + (\xi_2^{(\alpha)} + \xi_5^{(\alpha)} + \xi_6^{(\alpha)} + \xi_8^{(\alpha)}) \ln(1-\alpha_1) +$$
$$(\xi_1^{(\alpha)} + \xi_3^{(\alpha)} + \xi_4^{(\alpha)} + \xi_7^{(\alpha)}) \ln(1-\alpha_2) + (\xi_1^{(\alpha)} + \xi_2^{(\alpha)} + \xi_4^{(\alpha)} + \xi_8^{(\alpha)}) \ln(1-\alpha_3) =$$
$$\ln\left(\alpha_1^{\alpha_1}\alpha_2^{\alpha_2}\alpha_3^{\alpha_3}(1-\alpha_1)^{1-\alpha_1}(1-\alpha_2)^{1-\alpha_2}(1-\alpha_3)^{1-\alpha_3}\right).$$

■

5.3. The Entropy Function and its Second Derivative

Given k, $k = 1, \ldots, 8$, the Hessian of the function $E_\alpha^{(k)}(\theta)$, $\theta \in \mathring{I}_7(\alpha)$, is the 4×4 symmetric matrix $\mathcal{H}^{(k)}(\theta) = (\frac{\partial^2 E_\alpha^{(k)}}{\partial\theta_i\partial\theta_j}(\theta))_{i,j=1}^4$, where $\frac{\partial^2 E_\alpha^{(k)}}{\partial\theta_i\partial\theta_j}(\theta) = -\frac{\partial\xi_k(\theta)}{\partial\theta_i}\frac{\partial\xi_k(\theta)}{\partial\theta_j}\frac{1}{\xi_k(\theta)}$. Then the Hessian $\mathcal{H}(\theta)$ of the entropy function $E_\alpha(\theta)$ is the 4×4 symmetric matrix $\mathcal{H}(\theta) = \sum_{k=1}^8 \mathcal{H}^{(k)}(\theta)$. In accord with [4, Ch. 3, 3.1.4], since the functions $E_\alpha^{(k)}(\theta)$ are strictly concave, the corresponding quadratic forms ${}^t\tau\mathcal{H}^{(k)}(\theta)\tau$ are negative semi-definite: ${}^t\tau\mathcal{H}^{(k)}(\theta)\tau \leq 0$ for all $\tau \in \mathbb{R}^4$. In particular, the quadratic form ${}^t\tau\mathcal{H}(\theta)\tau = \sum_{k=1}^8 {}^t\tau\mathcal{H}^{(k)}(\theta)\tau$ is negative semi-definite. Moreover, since ${}^t\tau\mathcal{H}^{(5)}(\theta)\tau = -\frac{1}{\theta_0}\tau_1^2$, ${}^t\tau\mathcal{H}^{(6)}(\theta)\tau = -\frac{1}{\theta_1}\tau_2^2$, ${}^t\tau\mathcal{H}^{(7)}(\theta)\tau = -\frac{1}{\theta_2}\tau_3^2$, and ${}^t\tau\mathcal{H}^{(8)}(\theta)\tau = -\frac{1}{\theta_3}\tau_4^2$, the quadratic form ${}^t\tau\mathcal{H}(\theta)\tau$ is negative definite for any $\theta \in \mathring{I}_7(\alpha)$ and we obtain

Lemma 7. The set of local maximums of the entropy function $E_\alpha(\theta)$ coincides with the set of its critical points.

The compactness of the polytope $I_7(\alpha)$ yields that the extended entropy function $\hat{E}_\alpha(\theta)$ attains its absolute maximum and absolute minimum.

Theorem 3. The extended entropy function $\hat{E}_\alpha(\theta)$ has a unique absolute maximum attained at the point $\theta^{(\alpha)}$ from (10).

Proof. Lemma 6 and Lemma 7 yield that the entropy function $E_\alpha(\theta)$ and, therefore, also the extended entropy function $\hat{E}_\alpha(\theta)$, has a local maximum at the point $\theta^{(\alpha)}$. In accord with Lemma 4 and Lemma 10, $\hat{E}_\alpha(\theta)$ has a unique absolute maximum at $\theta^{(\alpha)}$.

■

Theorem 4. If the extended entropy function $\hat{E}_\alpha(\theta)$ attains an absolute minimum at some point from the polytope $I_7(\alpha)$, then this point is a vertex of $I_7(\alpha)$.

Proof. Lemma 2 allows us to use [2, Theorem 12.1.5, 12.1.8, Proposition 12.1.9] and we conclude that since the restriction of $\hat{E}_\alpha(\theta)$ on an i-face, $i = 1, 2, 3$, of the polytope $I_7(\alpha)$ is also a strictly concave function, we can apply at most four times Lemma 11.

■

The continuous extension $\hat{E}_\alpha(\theta)$, $\theta \in I_7(\alpha)$, of the entropy function $E_\alpha(\theta)$, $\theta \in \mathring{I}_7(\alpha)$, is said to be the *extended entropy function of Yule's triples of type* $[(\alpha)]$.

6. Degree of Mutual Dependence of a Triple of Events

6.1. Two Motivation Statements

Lemma 8. The three components of the Yule's triple $A = (A_1, A_2, A_3)$ are mutually independent if and only if $\theta^{(A)} = \theta^{(\alpha)}$.

Proof. In accord with [8, I,§5, (4)], the events A_1, A_2, A_3 are mutually independent if and only if $\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$, $1 \leq i < j \leq 3$, $\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1)\Pr(A_2)\Pr(A_3)$. Using (2), we write these conditions in the form

$$\left|\begin{array}{lllll} \theta_0 & + & \theta_1 & = & \alpha_2\alpha_3 \\ \theta_0 & + & \theta_2 & = & \alpha_1\alpha_3 \\ \theta_0 & + & \theta_3 & = & \alpha_1\alpha_2 \\ \theta_0 & & & = & \alpha_1\alpha_2\alpha_3. \end{array}\right.$$

The point $\theta^{(\alpha)}$ from (10) is the unique solution of this system.

■

Now, we suppose, in addition, that $(\Omega, \mathcal{A}, \Pr)$ is a discrete uniform probability space. The faces of the polytope $I_7(\alpha) \subset \mathbb{R}^4$ are parts of the hyperplanes with equations $\xi_k(\theta) = 0$, $k = 1, \ldots, 8$. According to (1), the following equivalences hold:

Lemma 9. Let $A = (A_1, A_2, A_3)$ be a Yule's triple of events. One has:

$$\xi_1(\theta^{(A)}) = 0 \text{ iff } A_1 \subset A_2 \cup A_3, \xi_2(\theta^{(A)}) = 0 \text{ iff } A_2 \subset A_1 \cup A_3,$$

$$\xi_3(\theta^{(A)}) = 0 \text{ iff } A_3 \subset A_1 \cup A_2, \xi_4(\theta^{(A)}) = 0 \text{ iff } A_1^c \subset A_2 \cup A_3,$$

$$\xi_5(\theta^{(A)}) = 0 \text{ iff } A_1 \cap A_2 \subset A_3^c, \xi_6(\theta^{(A)}) = 0 \text{ iff } A_2 \cap A_3 \subset A_1,$$

$$\xi_7(\theta^{(A)}) = 0 \text{ iff } A_1 \cap A_3 \subset A_2, \xi_8(\theta^{(A)}) = 0 \text{ iff } A_1 \cap A_2 \subset A_3.$$

6.2. Definition of Degree of Mutual Dependence

The value of extended entropy function $\hat{E}_\alpha(\theta)$ of Yule's triples of type $[(\alpha)]$ at $\theta = \theta^{(A)}$ is called *entropy of Yule's triple* $A = (A_1, A_2, A_3)$ *of type* $[(\alpha)]$. In accord with Corollary 2, the entropy does not depend on the order of the components of A. This fact together with the opposites described in Lemmas 8 and 9 motivate the use of the extended entropy function $\hat{E}_\alpha(\theta)$ as a measure of strength of mutual dependence of three events A_1, A_2, A_3.

Let us denote by M the absolute maximum $\hat{E}_\alpha(\theta^{(\alpha)})$ and let m be the absolute minimum of $\hat{E}_\alpha(\theta)$, attained at some vertex of the polytope $I_7(\alpha)$, see Theorems 3 and 4. The former also yields that $m < M$.

Following [6, 5.2], for any $\theta \in I_7(\alpha)$ we define $e_\alpha \colon I_7(\alpha) \to [0,1], e_\alpha(\theta) = \frac{\hat{E}_\alpha(\theta) - M}{m - M}$. The value of the function e_α at $\theta \in I_7(\alpha)$, $\theta = \theta^{(A)}$, $A = (A_1, A_2, A_3)$, is said to be *degree of mutual dependence of the events* A_1, A_2, A_3, *with* $\alpha_1 = \Pr(A_1)$, $\alpha_2 = \Pr(A_2)$, $\alpha_3 = \Pr(A_3)$. Intuitively, $e_\alpha(\theta^{(A)})$ measures the strength of the mutual relations among the events A_1, A_2, A_3.

The above definition of e_α yields

Corollary 3. The degree of mutual dependence of three events does not depend on the choice of base of logarithms in the extended entropy function.

Example 5. In case $\alpha = (\frac{1}{10}, \frac{1}{5}, \frac{3}{10})$ the polytope $I_7(\alpha)$ has 12 vertices

$$v_{1,2,3,8}, v_{1,2,5,8}, v_{1,3,5,8}, v_{2,3,5,8}, v_{1,2,3,5}, v_{1,2,5,7},$$

$$v_{1,2,7,8}, v_{1,5,6,7}, v_{1,5,6,8}, v_{1,6,7,8}, v_{2,5,7,8}, v_{5,6,7,8}.$$

Here by v_{k_1,k_2,k_3,k_4} we denote the vertex which is the intersection point of the hyperplanes with equations $\xi_{k_1} = 0$, $\xi_{k_2} = 0$, $\xi_{k_3} = 0$, and $\xi_{k_4} = 0$. At the first four vertices the extended entropy function attains its absolute minimum (approximately equal to 0.8018185525433372). Equivalently, we have

$$e_\alpha(v_{1,2,3,8}) = e_\alpha(v_{1,2,5,8}) = e_\alpha(v_{1,3,5,8}) = e_\alpha(v_{2,3,5,8}) = 1.$$

On the other hand, let, for example, the vertex $v_{1,3,5,8}$ belongs to the dotted polytope $I_7^{(\cdot)}(\alpha)$, that is, let $\theta^{(A)} = v_{1,3,5,8}$, where $A = (A_1, A_2, A_3)$ is a Yule's triple.

Moreover, let us assume that $(\Omega, \mathcal{A}, \Pr)$ is a sample space with equally likely outcomes. In accord with Lemma 9, we can conclude that the system of set-theoretic relations

$$A_1 \subset A_2 \cup A_3, A_3 \subset A_1 \cup A_2, A_1 \cap A_2 \subset A_3^c, A_1 \cap A_2 \subset A_3,$$

or equivalently, the system of relations $A_3 \subset A_1 \cup A_2$, $A_1 \subset A_3 \cap A_2^c$, is one of the most powerful under the condition $\alpha = (\frac{1}{10}, \frac{1}{5}, \frac{3}{10})$.

On the other hand, $v_{1,3,5,8}$ is again a vertex in case $\alpha = (\frac{1}{5}, \frac{3}{10}, \frac{2}{5})$ but now the above system of relations is not the most powerful one: $e_\alpha(v_{1,3,5,8}) < 1$.

Example 6. [9, Section 3, 3.2], (Bernstein 1928) Let us consider a sample space with four equally likely outcomes 112, 121, 211, 222. The events $A_1 = \{112, 121\}, A_2 = \{112, 211\}, A_3 = \{121, 211\}$, are pairwise independent but not mutually independent because $A_1 \cap A_2 \cap A_3 = \emptyset$. Below we evaluate their degree of mutual dependence. We set $A = (A_1, A_2, A_3)$ and note that $\alpha = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Using (1), we obtain $\xi_1^{(A)} = \xi_2^{(A)} = \xi_3^{(A)} = \xi_5^{(A)} = 0$, $\xi_4^{(A)} = \xi_6^{(A)} = \xi_7^{(A)} = \xi_8^{(A)} = \frac{1}{4}$. Therefore $\hat{E}_\alpha(\theta^{(A)}) = -2\ln\frac{1}{2}$. On the other hand, the polytope $I_7(\alpha)$ has 50 vertices and the extended entropy function $\hat{E}_\alpha(\theta)$ attains its absolute minimum $m = -\ln\frac{1}{2}$ at 48 of them. Since $M = \hat{E}_\alpha(\xi^{(\alpha)}) = -3\ln\frac{1}{2}$, we have $e_\alpha(\theta^{(A)}) = \frac{1}{2}$.

Remark 1. One can find below the link to a Java program which calculates the degree of mutual dependence of three events in a sample space with equally likely outcomes:
`http://www.math.bas.bg/algebra/valentiniliev/`

7. Conclusions

This paper finishes the trilogy that begins with [6] and [7]. It presents an original approach to the problem of measuring the magnitude of dependence of several events in a probability space, which rests upon Boltzmann-Shannon entropy of a probability distributions produced by these events. The first two parts are devoted to the fundamental case of two events where, for a given level of entropy intensity, one can discern negative from positive dependence, thus defining a direction. Moreover, the function of dependence of two events is closely related to the information exchanged between the two binary trials generated by these events.

The case of three events is studied here and this examination shows, in particular, that the general case of a finite number of events differs only in technical difficulties.

A. Appendix

A.1. Folklore Results about Extrema of a Concave Function

Our source of definitions and results about convex sets is [1, Ch. 11].

Let $C \subset \mathbb{R}^n$. We remind that the function $f\colon C \to \mathbb{R}$ is said to be *concave* (respectively, *strictly concave*) if C is a convex set and for any two different points $c_1, c_2 \in C$ and any $\lambda \in (0,1)$ one has $f((1-\lambda)c_1 + \lambda c_2) \geq (1-\lambda)f(c_1) + \lambda f(c_2)$ (respectively, $f((1-\lambda)c_1 + \lambda c_2) > (1-\lambda)f(c_1) + \lambda f(c_2)$).

Lemma 10. (i) Any local maximum point of a concave function is an absolute one.

(ii) There exists at most one local maximum point of a strictly convex function.

(iii) There exists at most one absolute maximum point of a strictly concave function.

Proof. Let $f\colon C \to \mathbb{R}$ be a concave function.

(i) Let $c_0 \in C$ be a point at which f attains a local maximum and let $U \subset C$ be a neighbourhood of c_0 such that $f(c_0) \leq f(c)$ for all $c \in U$. Let us suppose that there exists a point $c_1 \in C$ such that $f(c_1) > f(c_0)$. Then $f((1-\lambda)c_0 + \lambda c_1) \leq (1-\lambda)f(c_0) + \lambda f(c_1) > f(c_0)$ for all $\lambda \in (0,1)$. If λ is sufficiently close to 0, then $f((1-\lambda)c_0 + \lambda c_1) \in U$ and hence $f((1-\lambda)c_0 + \lambda c_1) \geq f(c_0)$ which is a contradiction.

(ii) Let, in addition, f be strictly concave and $c_1, c_2 \in C$ be two different points at which f attains a local maximum. In accord with part (i), we have $f(c_1) = f(c_2)$ and then $f((1-\lambda)c_1 + \lambda c_2) > (1-\lambda)f(c_1) + \lambda f(c_2) = f(c_1)$ for all $\lambda \in (0,1)$. Since f attains an absolute maximum at c_1, this is a contradiction.

Part (ii) implies part (iii).

■

Lemma 11. Let $f\colon C \to \mathbb{R}$ be a strictly concave function and let for any point $c \in \mathring{C}$ there exists an open line segment W_c such that $c \in W_c \subset C$. If f attains an absolute minimum at $c_0 \in C$, then $c_0 \notin \mathring{C}$.

Proof. Let us suppose that $c_0 \in \mathring{C}$ and let the points $c_1, c_2 \in W_c$, $c_1 \neq c_2$, be such that $c_0 = (1-\lambda)c_1 + \lambda c_2$ for some $\lambda \in (0,1)$. Then $f(c_1) \geq f(c_0)$, $f(c_2) \geq f(c_0)$, and $f(c_0) = f((1-\lambda)c_1 + \lambda c_2) > (1-\lambda)f(c_1) + \lambda f(c_2) \geq (1-\lambda)f(c_0) + \lambda f(c_0) = f(c_0)$, which is a contradiction. ■

Acknowledgements

It is a pleasure for me to cordially thank Dimitar Guelev for making experimental implementation of the evaluation of degree of dependence of three events from a discrete uniform probability space in Java. His numerical examples were invaluable for my work. I would like to thank also to the administration of the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences for creating perfect and safe conditions of work.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of Conflicting Interests

The Author declares that there is no conflict of interest.

References

[1] Berger M. (1987) Geometry I, Springer-Verlag Berlin Heidelberg.
[2] Berger M. (1987) Geometry II, Springer-Verlag Berlin Heidelberg.
[3] Bourbaki N. (1966) General Topology, Hermann, Paris.
[4] Boyd S, Vandenberghe L. (2004) Convex Optimization, Cambridge University Press.
[5] Husemoller D. (1966) Fibre Bundles, McGraw-Hill.
[6] Iliev V. V. (2021) On the Use of Entropy as a Measure of Dependence of Two Events. *Reliability: Theory & Applicatins* 16:237–248.
[7] Iliev V. V. (2022) On the Use of Entropy as a Measure of Dependence of Two Events. Part 2. *Reliability: Theory & Applicatins* 17:441–446.
[8] Kolmogorov A. N. (1956). Foundations of the Theory of Probability, Chelsea Publishing Company, New Yourk.
[9] Stoyanov J .M. (2013) Counterexamples in Probability, Dover Publications, Inc., Mineola, New York.