On the Use of Entropy as a Measure of Dependence of Two Events

Valentin Vankov Iliev

Institute of Mathematics and Informatics

Bulgarian Academy of Sciences

Sofia, Bulgaria

e-mail: viliev@math.bas.bg

Abstract

We define degree of dependence of two events A and B in a probability space by using Boltzmann-Shannon entropy function of an appropriate probability distribution produced by these events and depending on one parameter (the probability of intersection of A and B) varying within a closed interval I.

The entropy function attains its global maximum when the events A and B are independent. The important particular case of discrete uniform probability space motivates this definition in the following way. The entropy function has a minimum at the left endpoint of I exactly when one of the events and the complement of the other are connected with the relation of inclusion (maximal negative dependence). It has a minimum at the right endpoint of I exactly when one of these events is included in the other (maximal positive dependence).

Moreover, the deviation of the entropy from its maximum is equal to average information that carries one of the binary trials $A \cup A^c$ and $B \cup B^c$ with respect to the other. As a consequence, the degree of dependence of A and B can be expressed in terms of information theory and is invariant with respect to the choice of unit of information.

Using this formalism, we describe completely the screening tests and their reliability, measure efficacy of a vaccination, the impact of some events from the financial markets to other events, etc.

A link is available for downloading an Excel program which calculates the degree of dependence of two events in 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, Definitions, Notation

1.1   Introduction

In this paper we study the set of ordered pairs (A, B) of events in a probability space in order to define a measure of dependence (the power of relations) between A and B. This is done by means of Boltzmann-Shannon entropy of a variable probability distribution that arises naturally out of the pair (A, B). This approach is radically different from the standard ones where most of the measures of dependence are linear functions in the probability of intersection $A \cap B$, see Section 4. For a detailed study of these measures refer to [1].

The ordered pairs (A, B) in the form of two-attributed tables are used meticulously by G. Udny Yule in his memoir [9] in order to "...classify the objects or individuals observed into two classes only". He presents various examples and defines different indices to study the degree of association (dependence) of the corresponding two events. Any such ordered pair of events is said to be, as we termed it, an *Yule's pair of events*. G. Udny Yule himself noted that W. R. Macdonell in [4] used a two-attributed table as a tool to study "...the degree of effectiveness of vaccination in small-pox".

The paper is organized as follows. In Section 2 we parameterize the members of an equivalence class consisting of Yule's pairs with fixed probabilities α and β (that is, Yule's pairs of type (α, β)). We use the fact that the probability distribution produced by the probabilities of results of the experiment corresponding to a Yule's pair (cf. [3, I,§5]) is solution of a linear system with one free variable θ (see (2.1.3)). The system of inequalities that restrict the components of this solution is equivalent to the restriction of the variation of θ within a closed interval $I(\alpha, \beta) \subset [0, 1]$. Thus, we naturally introduce

(α, β, θ)-equivalence classes of Yule's pairs, whose members are said to be Yule's pairs of type (α, β, θ). Note that for any such pair, θ is the probability of the intersection of its components.

When we vary $(\alpha, \beta) \in [0, 1]^2$, the segment $\{\alpha\} \times \{\beta\} \times I(\alpha, \beta)$ sweeps a tetrahedron T_3 in $\mathbb{R}^3$, so the (α, β, θ)-equivalence classes of Yule's pairs are represented by some points in T_3, which, in turn, form so called dotted tetrahedron $T_3^{(\cdot)}$.

On the other hand, the affine isomorphism (2.1.4) which transforms $\mathbb{R}^3$ onto the hyperplane H in $\mathbb{R}^4$ that contains the solutions of the linear system (2.1.3), transforms the tetrahedron T_3 onto the 3-simplex $\Delta_3 \subset H$. Moreover, the dotted tetrahedron $T_3^{(\cdot)}$ is mapped on the dotted 3-simplex $\Delta_3^{(\cdot)} \subset H$, the latter classifying the probability distributions produced by all Yule's pairs. For the precise statements see Theorem 2.1.2 and Figure 1.

Given a Yule's pair of type (α, β, θ), Boltzmann-Shannon entropy $E_{\alpha, \beta}(\theta)$ of its distribution is a continuous function in $\theta \in I(\alpha, \beta)$ and its behaviour is described in Theorem 3.1.1 from Section 3. In particular, we show that $E_{\alpha, \beta}(\theta)$ attains its global maximum at the only point $\theta_0 = \alpha\beta$ for which the components of all Yule's pairs of type $(\alpha, \beta, \theta_0)$ (if any) are independent. The special case of a sample space with equally likely outcomes illustrates the fact that the maximum of dependence occurs at the endpoints of the interval $I(\alpha, \beta)$. More precisely, at the left endpoint we have $A \subset B^c$ or $B^c \subset A$ and at the right endpoint — $A \subset B$ or $B \subset A$.

Finally, $E_{\alpha, \beta}(\theta) = E_{\beta, \alpha}(\theta)$ and this common entropy function strictly increases to the left of θ_0 an strictly decreases to the right.

All of this motivates the use of entropy function $E_{\alpha, \beta}(\theta)$ as a measure of dependence of two events A and B with $\Pr(A) = \alpha$ and $\Pr(B) = \beta$: Negative dependence to the left of θ_0 and positive dependence to the right. By modifying appropriately $E_{\alpha, \beta}(\theta)$ by linear functions, we obtain a strictly increasing continuous function $e_{\alpha, \beta}$ which maps the range of θ onto the interval $[-1, 1]$ and serves (and is termed) as degree of dependence of the events A and B.

From the link provided in Remark 3.2.1 one can download a simple Excel program which calculates the degree of dependence of two events in a discrete uniform probability space with given cardinalities and given cardinality of their intersection.

It turns out that the expression for entropy function $E_{\alpha, \beta}$ is a particular case of what Shannon called in [6, Part I, Sction 6] the entropy of the joint event.

More precisely, this is the complete amount of information which contains in the results of the experiment $\mathfrak{J}$ from (2.1.1). On the other hand, $\mathfrak{J}$ is the joint experiment of two binary trials: $\mathfrak{A} = A \cup A^c$ and $\mathfrak{B} = B \cup B^c$. Theorem 3.3.1 shows that the mutual information $I(\mathfrak{A}, \mathfrak{B})$ of the experiments $\mathfrak{A}$ and $\mathfrak{B}$ is equal to the deviation of the entropy $E_{\alpha,\beta}(\theta)$ from its maximum $E_{\alpha,\beta}(\alpha\beta)$. In accord with the expression (3.3.2) which represents the function $e_{\alpha,\beta}(\theta)$ as a fraction of amounts of information, the degree of dependence of two events is invariant with respect to change of unit of information (bits, nats, etc.).

In case Yule's pairs are models of a screening tests, the probability $F_-(\theta)$ of false negative and the probability $F_+(\theta)$ of false positive test are tending from statistically insignificant nearby the left endpoint of the range of θ to statistical significance in a neighbourhood of the right endpoint. Moreover, on the complement of any such neighbourhood the product $F_-(\theta)F_+(\theta)$ is bound below by a positive constant. In other words, a kind of uncertainty principle holds — see Subsection 3.4.

In Subsection 3.5 we show that the degree of dependence of pairs of events can be used as a measure of effectiveness of vaccine for a particular decease. As an example we estimate the efficacy of vaccine for small-pox tested via the epidemic at Sheffield in 1887-88, the statistical data taken from [9, I] .

In Section 4 we give several examples of other measures of dependence which are evaluated by using Sheffield's sample.

1.2   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:

R is the range of the probability function $\Pr \colon \mathcal{A} \to \mathbb{R}$; $[(\alpha, \beta)]$ is the fiber of the surjective map $\mathcal{A}^2 \to R^2, (A, B) \mapsto (\Pr(A), \Pr(B))$, over $(\alpha, \beta) \in R^2$; $\theta^{(A,B)} = \Pr(A \cap B)$, $(A, B) \in \mathcal{A}^2$; $[(\alpha, \beta, \theta)]$ is the fiber of the map $[(\alpha, \beta)] \to R$, $(A, B) \mapsto \theta^{(A,B)}$, with image $R^{(\alpha,\beta)} \subset R$, over any $\theta \in R^{(\alpha,\beta)}$.

We note that the fibers $[(\alpha, \beta)]$ for $(\alpha, \beta) \in R^2$ form a partition of $\mathcal{A}^2$ and the fibers $[(\alpha, \beta, \theta)]$ for $\theta \in R^{(\alpha,\beta)}$ form a partition of $[(\alpha, \beta)]$.

As usual, the events $\emptyset$ and Ω are called *trivial*. The members of the equiv-

alence class $[(\alpha, \beta)]$ (resp., the equivalence class $[(\alpha, \beta, \theta)]$) are said to be *Yule's pairs of type* (α, β) (resp., *Yule's pairs of type* (α, β, θ)).

2 Classification of Yule's Pairs

2.1 The Probability Distribution of a Yule's Pair

Any ordered pair $(A, B) \in \mathcal{A}^2$ produces an experiment

$$\mathfrak{J} = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B) \cup (A^c \cap B^c) \tag{2.1.1}$$

(cf. [3, I,§5]) and the probabilities of its results:

$$\xi_1^{(A,B)} = \Pr(A \cap B), \xi_2^{(A,B)} = \Pr(A \cap B^c),$$

$$\xi_3^{(A,B)} = \Pr(A^c \cap B), \xi_4^{(A,B)} = \Pr(A^c \cap B^c).$$

For any $(A, B) \in [(\alpha, \beta)]$, the probability distribution

$$(\xi_1, \xi_2, \xi_3, \xi_4) = (\xi_1^{(A,B)}, \xi_2^{(A,B)}, \xi_3^{(A,B)}, \xi_4^{(A,B)}) \tag{2.1.2}$$

satisfies the linear system

$$\begin{vmatrix} \xi_1 & + & \xi_2 & & & & & = & \alpha \\ & & & & \xi_3 & + & \xi_4 & = & 1 - \alpha \\ \xi_1 & & & + & \xi_3 & & & = & \beta \\ & & \xi_2 & & & + & \xi_4 & = & 1 - \beta. \end{vmatrix} \tag{2.1.3}$$

Let H be the affine hyperplane in $\mathbb{R}^4$ with equation $\xi_1 + \xi_2 + \xi_3 + \xi_4 = 1$. The solutions of (2.1.3) depend on one parameter, say $\theta = \xi_1$, and form a straight line $\ell_{\alpha,\beta}$ in H with parametric representation

$$\ell_{\alpha,\beta} \colon \xi_1 = \theta, \xi_2 = \alpha - \theta, \xi_3 = \beta - \theta, \xi_4 = 1 - \alpha - \beta + \theta.$$

The map

$$\iota \colon \mathbb{R}^3 \to H, (\alpha, \beta; \theta) \mapsto (\theta, \alpha - \theta, \beta - \theta, 1 - \alpha - \beta + \theta) \tag{2.1.4}$$

is an affine isomorphism with inverse affine isomorphism

$$\chi \colon H \to \mathbb{R}^3, \xi \mapsto (\xi_1 + \xi_2, \xi_1 + \xi_3, \xi_1).$$

The trace of the 4-dimensional cube $\{\xi \in \mathbb{R}^4 | 0 \leq \xi_k \leq 1, k = 1, 2, 3, 4\}$ onto the hyperplane H is the 3-dimensional simplex, that is, the tetrahedron, Δ_3 defined in H by the inequalities

$$\xi_1 \geq 0, \xi_2 \geq 0, \xi_3 \geq 0, \xi_1 + \xi_2 + \xi_3 \leq 1.$$

The inverse image $T_3 = \iota^{-1}(\Delta_3)$ via the affine isomorphism ι is the tetrahedron in $\mathbb{R}^3$ defined by the system of inequalities

$$\theta \leq \alpha, \theta \leq \beta, \theta \geq \alpha + \beta - 1, \theta \geq 0. \qquad (2.1.5)$$

In other words, this is the tetrahedron with vertices $O(0,0,0)$, $M(1,0,0)$, $N(0,1,0)$, $P(1,1,1)$ — see Figure 1.

For any $(\alpha, \beta) \in \mathbb{R}^2$ we set $C(\alpha, \beta) = \lambda_{\alpha,\beta} \cap T_3$, so $C(\alpha, \beta) = \{\alpha\} \times \{\beta\} \times I(\alpha, \beta)$, $I(\alpha, \beta) \subset \mathbb{R}$. The system (2.1.5) yields that $I(\alpha, \beta)$ equals the closed interval $[\ell(\alpha, \beta), r(\alpha, \beta)]$, $\ell(\alpha, \beta) = \max(0, \alpha + \beta - 1)$, $r(\alpha, \beta) = \min(\alpha, \beta)$. We have $\alpha\beta \in I(\alpha, \beta)$ and denote by $\mathring{I}(\alpha, \beta)$ the interior of the interval $I(\alpha, \beta)$. We obtain immediately:

Lemma 2.1.1 *Let $(\alpha, \beta) \in [0,1]^2$. The next three statements are equivalent:*

(i) *One has $(\alpha, \beta) \in (0,1)^2$.*

(ii) *One has $\alpha\beta \in \mathring{I}(\alpha, \beta)$.*

(iii) *One has $\mathring{I}(\alpha, \beta) \neq \emptyset$.*

(iv) *Under the above conditions, one has $\xi_k(\theta) > 0$ for all $\theta \in \mathring{I}(\alpha, \beta)$ and all $k = 1, 2, 3, 4$.*

(v) *Conversely, if there exists $\theta \in I(\alpha, \beta)$ such that $\xi_k(\theta) > 0$ for all $k = 1, 2, 3, 4$, then* (i) — (iii) *hold.*

We have $R^{(\alpha,\beta)} \subset I(\alpha, \beta)$ and define the *dotted interval* $I^{(\cdot)}(\alpha, \beta) = R^{(\alpha,\beta)}$. The *dotted segment* $C^{(\cdot)}(\alpha, \beta) = \{\alpha\} \times \{\beta\} \times I^{(\cdot)}(\alpha, \beta)$, $(\alpha, \beta) \in R^2$, is the locus of all triples of probabilities $(\alpha, \beta, \theta^{(A,B)})$, where $(A, B) \in [(\alpha, \beta)]$.

For any $(\alpha, \beta) \in \mathbb{R}^2$ we set $D(\alpha, \beta) = \iota(C(\alpha, \beta))$. Since $\iota(\lambda_{\alpha,\beta}) = \ell_{\alpha,\beta}$, we obtain that $D(\alpha, \beta) = \ell_{\alpha,\beta} \cap \Delta_3$.

Let $I_k(\alpha, \beta) = [\ell_k(\alpha, \beta), r_k(\alpha, \beta)]$ be the corresponding range of the real variable $\xi_k(\theta)$ for $k = 1, 2, 3, 4$, with $I_1(\alpha, \beta) = I(\alpha, \beta)$. The line segment $D(\alpha, \beta)$ in $\ell_{\alpha,\beta}$ has endpoints

$$(\ell_1(\alpha, \beta), \ell_2(\alpha, \beta), \ell_3(\alpha, \beta), \ell_4(\alpha, \beta)), (r_1(\alpha, \beta), r_2(\alpha, \beta), r_3(\alpha, \beta), r_4(\alpha, \beta)).$$

Since ι is also a homeomorphism, we have $\mathring{D}(\alpha, \beta) = \ell_{\alpha, \beta} \cap \mathring{\Delta}_3$.

The line segment $D(\alpha, \beta)$ contains the *dotted segment* $D^{(\cdot)}(\alpha, \beta)$ which is the locus of all probability distributions (2.1.2) for which $(A, B) \in [(\alpha, \beta)]$.

Finally, we note that $T_3 = \cup_{(\alpha,\beta) \in [0,1]^2} C(\alpha, \beta)$, $\Delta_3 = \cup_{(\alpha,\beta) \in [0,1]^2} D(\alpha, \beta)$, and the unions $T_3^{(\cdot)} = \cup_{(\alpha,\beta) \in R^2} C^{(\cdot)}(\alpha, \beta)$, $\Delta_3^{(\cdot)} = \cup_{(\alpha,\beta) \in R^2} D^{(\cdot)}(\alpha, \beta)$ are the corresponding *dotted tetrahedrons*.

The above considerations and Figure 1 yield the following theorem and its corollary.

Theorem 2.1.2 (i) *The affine isomorphism $\iota \colon \mathbb{R}^3 \to H$ from (2.1.4) is a strictly increasing transformation of any line segment $C(\alpha, \beta)$ (resp., dotted line segment $C^{(\cdot)}(\alpha, \beta)$) onto the line segment $D(\alpha, \beta)$ (resp., onto the dotted line segment $D^{(\cdot)}(\alpha, \beta)$).*

(ii) *ι maps the tetrahedron T_3 (resp., dotted tetrahedron $T_3^{(\cdot)}$) onto the tetrahedron Δ_3 (resp., onto the dotted tetrahedron $\Delta_3^{(\cdot)}$).*

(iii) *The dotted tetrahedron $T_3^{(\cdot)}$ is the classification space of all equivalence classes $[(\alpha, \beta, \theta)]$ of Yule's pairs.*

(iv) *The dotted tetrahedron $\Delta_3^{(\cdot)}$ is the classification space of all probability distributions (2.1.2) produced by Yule's pairs.*

Corollary 2.1.3 (i) *One has $\xi_1(\theta) = 0$ if and only if $(\alpha, \beta, \theta) \in MON$.*
(ii) *One has $\xi_2(\theta) = 0$ if and only if $(\alpha, \beta, \theta) \in NOP$.*
(iii) *One has $\xi_3(\theta) = 0$ if and only if $(\alpha, \beta, \theta) \in MOP$.*
(iv) *One has $\xi_4(\theta) = 0$ if and only if $(\alpha, \beta, \theta) \in MNP$.*

3   Entropy and Dependence of Yule's Pairs

3.1   Entropy of a Yule's Pair

Let us suppose that $(\alpha, \beta) \in (0, 1)^2$. Then Lemma 2.1.1 implies $\mathring{I}(\alpha, \beta) \neq \emptyset$ and $\xi_k(\theta) > 0$ for $\theta \in \mathring{I}(\alpha, \beta)$ and for all $k = 1, 2, 3, 4$. Therefore Boltzmann-Shannon entropy of the probability distribution $(\xi_1(\theta), \xi_2(\theta), \xi_3(\theta), \xi_4(\theta))$ is

defined (cf. [6], [7]):

$$E_{\alpha,\beta}(\theta) = -\sum_{k=1}^{4} \xi_k(\theta) \ln(\xi_k(\theta)), \theta \in \mathring{I}(\alpha,\beta).$$

Theorem 3.1.1 *Let* $(\alpha,\beta) \in (0,1)^2$.

(i) *For any* $\theta \in \mathring{I}(\alpha,\beta)$ *one has*

$$E'_{\alpha,\beta}(\theta) = \ln \frac{\xi_2(\theta)\xi_3(\theta)}{\xi_1(\theta)\xi_4(\theta)}.$$

(ii) *The function* $E_{\alpha,\beta}(\theta)$ *in* θ *strictly increases on the interval* $(\ell(\alpha,\beta),\alpha\beta]$ *and strictly decreases on the interval* $[\alpha\beta, r(\alpha,\beta))$, *having a global maximum at* $\theta = \alpha\beta$.

(iii) *The function* $E_{\alpha,\beta}(\theta)$ *can be extended uniquely as a continuous function on the closed interval* $I(\alpha,\beta)$, *which strictly increases on the interval* $[\ell(\alpha,\beta),\alpha\beta]$ *and strictly decreases on the interval* $[\alpha\beta, r(\alpha,\beta)]$.

(iv) *One has* $I(\alpha,\beta) = I(\beta,\alpha)$ *and* $E_{\alpha,\beta} = E_{\beta,\alpha}$.

Proof: (i) We have

$$E'_{\alpha,\beta}(\theta) = -\sum_{k=1}^{4} \xi'_k(\theta) \ln(\xi_k(\theta)) - \sum_{k=1}^{4} \xi_k(\theta) \frac{\xi'_k(\theta)}{\xi_k(\theta)} = \ln \frac{\xi_2(\theta)\xi_3(\theta)}{\xi_1(\theta)\xi_4(\theta)}.$$

(ii) The equation $E'_{\alpha,\beta}(\theta) = 0$ (resp., the inequality $E'_{\alpha,\beta}(\theta) > 0$) is equivalent to $\theta = \alpha\beta$ (resp., $\theta < \alpha\beta$).

(iii) According to Corollary 2.1.3, one or two functions $\xi_k(\theta)$ are zero at any endpoint of each interval $I(\alpha,\beta)$ and all functions $\xi_k(\theta)$ are strictly positive on the interior $\mathring{I}(\alpha,\beta)$. For a fixed interval $I(\alpha,\beta)$ and a fixed endpoint a of $I(\alpha,\beta)$ the limit $\lim_{\theta \to a} E_{\alpha,\beta}(\theta)$ exists and we extend $E_{\alpha,\beta}(\theta)$ as continuous at the point a.

(iv) We have $I(\alpha,\beta) = I(\beta,\alpha)$ and the transposition of events A and B yields transposition of the functions $\xi_2(\theta)$ and $\xi_3(\theta)$.

The continuous function $E_{\alpha,\beta}(\theta)$ in $\theta \in I(\alpha,\beta)$ is said to be the *entropy function of Yule's pairs of type* (α,β) and its value at $\theta = \theta^{(A,B)}$ is called *entropy of Yule's pair* (A,B) *of type* (α,β).

Theorem 3.1.1 implies immediately

Corollary 3.1.2 *Let $(\alpha, \beta) \in (0,1)^2$. The following three statements are equivalent:*

(i) *One has $\theta_0 = \alpha\beta$.*

(ii) *If Yule's pair (A, B) of type (α, β) satisfies the equality $\xi_1^{(A,B)} = \theta_0$, then the events A and B are independent.*

(iii) *The entropy function $E_{\alpha,\beta}(\theta)$ of Yule's pairs of type (α, β) attains its global maximum at the point θ_0.*

3.2   Degree of Dependence of Pairs of Events

We "normalize" the entropy function by composing the functions $E_{\alpha,\beta}(\theta)$ and $2E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\theta)$ on the intervals of their increase by the appropriate linear functions and obtain for any pair $(\alpha, \beta) \in (0,1)^2$ a continuous function $e_{\alpha,\beta} \colon I(\alpha, \beta) \to [-1, 1]$. In accord with Theorem 3.1.1, (iv), we have $e_{\alpha,\beta} = e_{\beta,\alpha}$. The value of the function $e_{\alpha,\beta}$ at $\theta \in I(\alpha, \beta)$, $\theta = \theta^{(A,B)}$, is said to be *degree of dependence of the events A and B with $\alpha = \Pr(A)$, $\beta = \Pr(B)$.*

The function $e_{\alpha,\beta}$ strictly increases on the interval $I(\alpha, \beta)$ from -1 to 1 and attains value 0 at the point $\alpha\beta$. The events A and B are said to be *negatively dependent* if $\theta^{(A,B)} < \alpha\beta$ and *positively dependent* if $\theta^{(A,B)} > \alpha\beta$. When $\theta^{(A,B)} = \alpha\beta$ the events A and B are independent (the entropy is maximal). In a small neighbourhood of the left endpoint $\ell(\alpha, \beta)$ of the interval $I(\alpha, \beta)$ the dependence is *negatively strong*, with "maximum" $1 = |-1|$ at $\theta = \ell(\alpha, \beta)$ (if attainable). In a small neighbourhood of the right endpoint $r(\alpha, \beta)$ of the interval $I(\alpha, \beta)$ the dependence is *positively strong*, with maximum 1 at $\theta = r(\alpha, \beta)$ (if attainable). In both cases, the entropy is minimal at the endpoints $\ell(\alpha, \beta)$ and $r(\alpha, \beta)$ of the corresponding semi-intervals. Note that in a small neighbourhood of the point $\theta = \alpha\beta$ the events A and B are "almost independent" (the entropy is close to its maximum).

Remark 3.2.1 One can find below the link to a simple Excel program which calculates the degree of dependence of two events in a sample space with equally likely outcomes:

`http://www.math.bas.bg/algebra/valentiniliev/`

3.3   A Glance at the Information Theory

The experiment $\mathfrak{J}$ from (2.1.1) is the joint experiment (see [6, Part I, Section 6]) of two simpler binary trials: $\mathfrak{A} = A \cup A^c$ and $\mathfrak{B} = B \cup B^c$ with $\Pr(A) = \alpha$, $\Pr(B) = \beta$. The *average quantity of information of one of the experiments* $\mathfrak{A}$ *and* $\mathfrak{B}$*, relative to the other,* (see [2, §1]), is defined in this particular case by the formula

$$I(\mathfrak{A}, \mathfrak{B}) = \xi_1 \ln \frac{\xi_1}{\alpha\beta} + \xi_2 \ln \frac{\xi_2}{\alpha(1-\beta)} + \xi_3 \ln \frac{\xi_3}{(1-\alpha)\beta} + \xi_4 \ln \frac{\xi_4}{(1-\alpha)(1-\beta)}.$$

$$(3.3.1)$$

The definition of the degree function $e_{\alpha,\beta}(\theta)$ and (3.3.1) yield immediately the following:

Theorem 3.3.1 (i) *One has*

$$I(\mathfrak{A}, \mathfrak{B})(\theta) = E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\theta)$$

for all $\theta \in I(\alpha, \beta)$.

(ii) *One has*

$$e_{\alpha,\beta}(\theta) = \begin{cases} -\dfrac{E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\theta)}{E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\ell(\alpha,\beta))} & \text{if } \ell(\alpha,\beta) \le \theta \le \alpha\beta \\ \dfrac{E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\theta)}{E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(r(\alpha,\beta))} & \text{if } \alpha\beta \le \theta \le r(\alpha,\beta). \end{cases}$$

$$(3.3.2)$$

Remark 3.3.2 Since

$$E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\ell(\alpha,\beta)) = \max_{\ell(\alpha,\beta) \le \tau \le \alpha\beta} I(\mathfrak{A}, \mathfrak{B})(\tau),$$

$$E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(r(\alpha,\beta)) = \max_{\alpha\beta \le \tau \le r(\alpha,\beta)} I(\mathfrak{A}, \mathfrak{B})(\tau),$$

we can write down the equality (3.3.2) in the form

$$e_{\alpha,\beta}(\theta) = \begin{cases} -\dfrac{I(\mathfrak{A},\mathfrak{B})(\theta)}{\max_{\ell(\alpha,\beta) \le \tau \le \alpha\beta} I(\mathfrak{A},\mathfrak{B})(\tau)} & \text{if } \ell(\alpha,\beta) \le \theta \le \alpha\beta \\ \dfrac{I(\mathfrak{A},\mathfrak{B})(\theta)}{\max_{\alpha\beta \le \tau \le r(\alpha,\beta)} I(\mathfrak{A},\mathfrak{B})(\tau)} & \text{if } \alpha\beta \le \theta \le r(\alpha,\beta). \end{cases}$$

Part (ii) of the above theorem implies

Corollary 3.3.3 *The degree of dependence of two events does not depend on the choice of unit of information.*

The graphs of $E_{\alpha,\beta}$ and $e_{\alpha,\beta}$ for some particular $(\alpha, \beta) \in (0,1)^2$ are presented in Figures 2 and 3.

3.4　Application: Description of a Screening Test

According to Merriam-Webster Dictionary, a screening test is "...a prelimi-
nary or abridged test intended to eliminate the less probable members of an
experimental series". In other words, some of the equally likely outcomes
of a sample space Ω possess a property and, in this way, form an event A.
On the other hand, there exists an event B consisting of all outcomes which
as if have this property after conducting the test. Thus, we obtain a Yule's
pair of events in a sample space Ω. The test does not always work perfectly
— sometimes it is negative under the condition that the property is present
(that is, *false negative*), and sometimes it is positive under the condition
that the property is absent (that is, *false positive*). Let us suppose that all
members of the population are tested and that $\Pr(A) = \alpha$, $\Pr(B) = \beta$, where
$(\alpha, \beta) \in (0, 1)^2$. Yule's pair (A, B) produces the experiment (2.1.1) and in
turn, the probability distribution (2.1.2) consisting of its results. In the nota-
tion introduced in Subsection 2.1, the probability F_- of false negative result
is $\frac{\xi_2^{(A,B)}}{\alpha} = \frac{\alpha-\theta}{\alpha}$ and the probability F_+ of false positive result is $\frac{\xi_3^{(A,B)}}{1-\alpha} = \frac{\beta-\theta}{1-\alpha}$,
where $\theta = \theta^{(A,B)}$. The product $F_-(\theta)F_+(\theta)$ is a quadratic function in θ which
strictly decreases on the interval $I(\alpha, \beta)$ and assumes value 0 at its right end-
point $r(\alpha, \beta)$. In particular, for the complement of any open neighbourhood
of $r(\alpha, \beta)$ in the interval $I(\alpha, \beta)$, there exists a positive constant K such that
$F_-(\theta)F_+(\theta) \geq K$ for any point θ from this complement. In other words,
both $F_-(\theta)$ and $F_+(\theta)$ can not be simultaneously as small as we want (a kind
of uncertainty principle). The conditional probabilities $F_-(\theta)$ and $F_+(\theta)$ are
statistically acceptable in a small neighbourhood of the point $r(\alpha, \beta)$, at least
one being 0 at this point. The reliability of $F_-(\theta)$ and $F_+(\theta)$ decreases when
θ approaches the left endpoint $\ell(\alpha, \beta)$ of $I(\alpha, \beta)$. When $\theta = \ell(\alpha, \beta)$, at least
one of $F_-(\theta)$ and $F_+(\theta)$ is equal to 1. In terms of the degree of dependence
$e_{\alpha,\beta}(\theta)$, $\theta = \theta^{(A,B)}$, of the events A and B, this behaviour can be described
in the following way: When $e_{\alpha,\beta}(\theta)$ is close to -1, then the test is not re-
liable but its effectiveness increases when $e_{\alpha,\beta}(\theta)$ approaches 1. In a small
neighbourhood of 1 the test is statistically acceptable.

3.5　Application: Effectiveness of Vaccination

Let us consider a population whose members have a particular disease for
which a vaccine is developed. Let A be the set of all those who have recovered

and let B be the set of vaccinated members. Then A^c is the set of all fatal endings and B^c is the set of unvaccinated. If $\alpha = \Pr(A)$, $\beta = \Pr(B)$, then the degree of dependence $e_{\alpha,\beta}(\theta)$, $\theta = \theta^{(A,B)}$, of the events A and B measures the effectiveness of the corresponding vaccine. More precisely, when $e_{\alpha,\beta}(\theta)$ is close to -1, then the vaccine is counterproductive and its effectiveness increases being negative when $e_{\alpha,\beta}(\theta) < 0$ and positive when $e_{\alpha,\beta}(\theta) > 0$. In case $e_{\alpha,\beta}(\theta) = 0$ the vaccination does not influence the recovery and it is very positively effective when $e_{\alpha,\beta}(\theta)$ is close to 1.

In his memoir [9, Section I, Table I], G. Udny Yule presents a table used by W. R. Macdonell in [4] in order to show "...the recoveries and deaths amongst vaccinated and unvaccinated patients during the small-pox epidemic at Sheffield in 1887-88", see Table 1.

We have $\alpha = 0.88262811$, $\beta = 0.899213268$, $\theta = 0.840102063$, and $e_{\alpha,\beta}(\theta) = 0.268810618$. Therefore the results of this vaccination are faintly positive (the recovery is not only due to vaccination!).

Yule's pair (A, B) considered as a screening test has statistically acceptable false negative probability $\Pr(B^c|A) \approx 0.0482$ (the probability that a member is unvaccinated under the condition that he/she is recovered). On the other hand, this test has not statistically significant false positive probability $\Pr(B|A^c) \approx 0.4964$ (the probability that a member of the population was vaccinated under the condition that he/she is dead). Equivalently: the matter of life and death depended of the result of tossing an almost fair coin!

4   Other Measures of Dependence

In this section we assume that Ω is a sample space with equally likely outcomes.

4.1   Yule's Q

The difference $\delta = \Pr(A \cap B) - \Pr(A)\Pr(B) = \theta - \alpha\beta$ (the deviation from independence) is called *copula between A and B*. It is cited by G. Udny Yule in [9, Section I, n_o 5]. He notes there that the relation $\delta = \xi_1(\theta)\xi_4(\theta) - \xi_2(\theta)\xi_3(\theta)$ is due to Karl Pearson (one of his teachers).

In [9, Section I, n_o 9], G. Udny Yule introduces his measure of association

first given in [8, Section I, n_o 9]:

$$Q = \frac{\xi_1(\theta)\xi_4(\theta) - \xi_2(\theta)\xi_3(\theta)}{\xi_1(\theta)\xi_4(\theta) + \xi_2(\theta)\xi_3(\theta)} = \frac{\theta - \alpha\beta}{2\theta^2 - (2\alpha + 2\beta - 1)\theta + \alpha\beta}.$$

It has the necessary properties: (a) $Q = 0$ if and only if A and B are independent; (b) $Q = 1$ if and only if $A \subset B$ or $B \subset A$; (c) $Q = -1$ if and only if $A \subset B^c$ or $B^c \subset A$. Finally, $-1 \leq Q \leq 1$. In the case of Sheffield's epidemic, we have $Q = 0.902299648$.

We define the function

$$Q_{\alpha,\beta}(\theta) = \frac{\theta - \alpha\beta}{2\theta^2 - (2\alpha + 2\beta - 1)\theta + \alpha\beta}, \theta \in I(\alpha, \beta),$$

which produces Yule's Q (see Figure 7).

Remark 4.1.1 There are infinitely many functions of the form $h(\theta) = \frac{f(\theta) - g(\theta)}{f(\theta) + g(\theta)}$ with the properties (a), (b), (c), and $-1 \leq h(\theta) \leq 1$, defined on the interval $I(\alpha, \beta)$, $(\alpha, \beta) \in (0, 1)^2$. For example, there exist infinitely many pairs of cubic polynomials $f(\theta)$ and $g(\theta)$ which work.

4.2 Obreshkoff's Measures of Dependence

The properties of the copula δ are also discussed by N. Obreshkoff in his textbook [5, Chapter 3,§6] and in [3]. In particular, the relation $\Pr(B|A) = \Pr(B) + \frac{\delta}{\Pr(A)}$ shows that "... the probability of one of these events increases under the condition that the other comes true in case $\delta > 0$ and decreases in case $\delta < 0$". Moreover, $-\delta = \Pr(A \cap B^c) - \Pr(A)\Pr(B^c)$.

The number

$$\rho(B; A) = \Pr(B|A) - \Pr(B|A^c) = \frac{\theta - \alpha\beta}{\alpha(1 - \alpha)}$$

is called *coefficient of regression of B with respect to A*. It measures the influence of A on B. We have $-1 \leq \rho(B; A) \leq 1$.

It has the following properties: (a) $\rho(B; A) = 0$ if and only if A and B are independent, (b) $\rho(B; A) = 1$ if and only if $A = B$, (c) $\rho(B; A) = -1$ if and only if $A^c = B$.

In the above example of small-pox epidemic at Sheffield we have $\rho(B; A) = 0.44819565$.

We define the functions

$$\rho_{\alpha;\beta}(\theta) = \frac{\theta - \alpha\beta}{\alpha(1 - \alpha)}, \rho_{\beta;\alpha}(\theta) = \frac{\theta - \alpha\beta}{\beta(1 - \beta)}, \theta \in I(\alpha, \beta),$$

which produce the corresponding coefficients of regression (see Figures 4 and 5).

The numbers $\rho(B; A)$ and $\rho(A; B)$ have the same sign and, in general, are not equal. Their geometric mean

$$R(A, B) = \pm\sqrt{\rho(A; B)\rho(B; A)} = \frac{\theta - \alpha\beta}{\sqrt{\alpha\beta(1 - \alpha)(1 - \beta)}},$$

where $\pm$ is chosen to be the common sign of $\rho(B; A)$ and $\rho(A; B)$, is said to be *coefficient of correlation between A and B*. This coefficient has the above properties (a) — (c). In the case of Sheffield's epidemic we have $R(A, B) = 0.4791876$.

We define the function

$$R_{\alpha;\beta}(\theta) = \frac{\theta - \alpha\beta}{\sqrt{\alpha\beta(1 - \alpha)(1 - \beta)}}, \theta \in I(\alpha, \beta),$$

which produces the corresponding coefficient of correlation (see Figure 6).

5   Conclusions

This paper presents an original approach to the problem of measuring the degree of dependence of two events A and B in a probability space. It uses the only reliable way of evaluation of the power of relations between these events, borrowed from statistical physics and information theory: this is the utilization of Boltzmann-Shannon entropy. More precisely, we start with the joint experiment assembled by the two binary trials $A \cup A^c$ and $B \cup B^c$. The four probabilities of results of this experiment constitute a variable probability distribution and satisfy a simple linear system whose general solution $(\xi_1, \xi_2, \xi_3, \xi_4)$ depends on one parameter θ (the probability of intersection $A \cap B$). Note that due to the natural constraints on ξ_k,

$k = 1, 2, 3, 4$, θ varies throughout a closed interval $I(\alpha, \beta)$, where $\alpha = \Pr(A)$ and $\beta = \Pr(B)$. We modify naturally the entropy function of the distribution $(\xi_1(\theta), \xi_2(\theta), \xi_3(\theta), \xi_4(\theta))$ and obtain the degree of dependence function $e_{\alpha,\beta}(\theta) \colon I(\alpha, \beta) \to [-1, 1]$. By definition, if $\theta = \Pr(A \cap B)$, then $e_{\alpha,\beta}(\theta)$ measures the intensity of relations between A and B.

Our degree of dependence is still within the probation period. In its defence it can be said that evaluates the mutual information which is exchanged between the random objects A and B and, moreover, does not depend on the choice of unit of information. It also reflects plausibly the behaviour of a screening test or impact of a vaccination on the survival of a person. The function $e_{\alpha,\beta}(\theta)$ can also be used for measuring the effectiveness of a drug or medical treatment, the association of adverse events with use of some particular drug, the association of certain events with the stock market prices, etc.

6 Figures and Tables
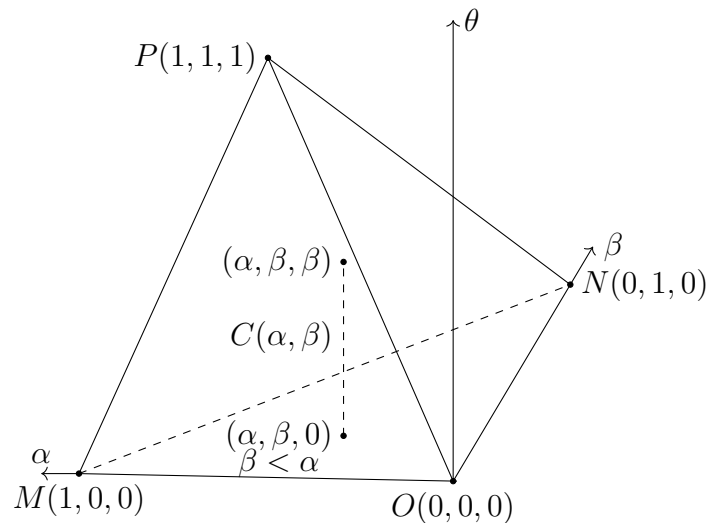
Figure 1

	B	B^c	Total
A	3951	200	4151
A^c	278	274	552
Total	4229	474	4703

Table 1

In all graphs below we use Sheffield's sample data.
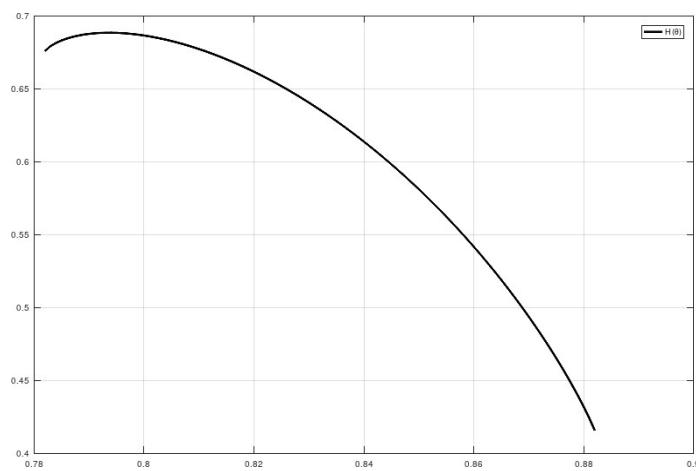
Figure 2

Graph of the entropy function $E_{\alpha,\beta}(\theta)$
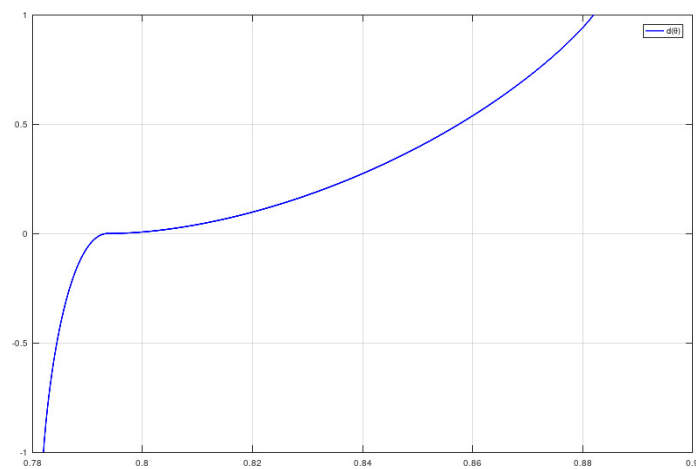
Figure 3

Graph of the degree function $e_{\alpha,\beta}(\theta)$
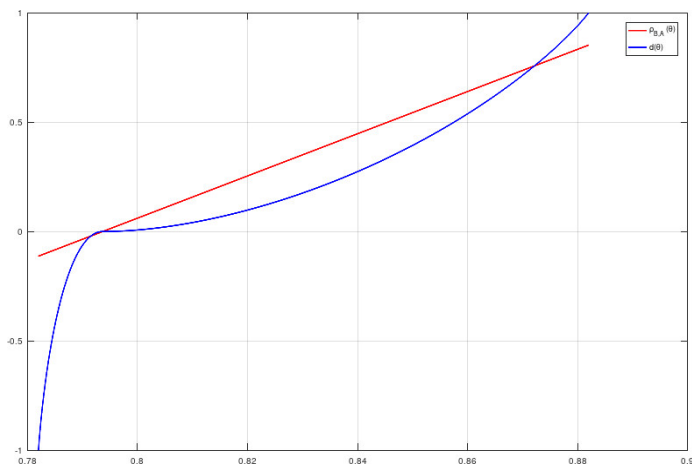
Figure 4

Comparison of the graphs of $e_{\alpha,\beta}(\theta)$ and $\rho_{\beta;\alpha}(\theta)$
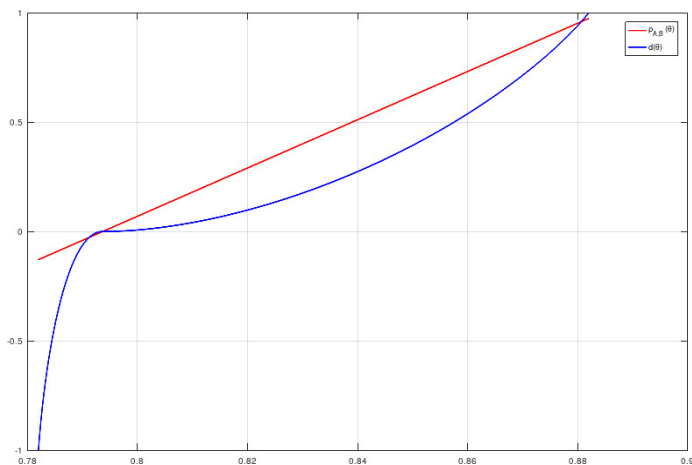
Figure 5

Comparison of the graphs of $e_{\alpha,\beta}(\theta)$ and $\rho_{\alpha;\beta}(\theta)$
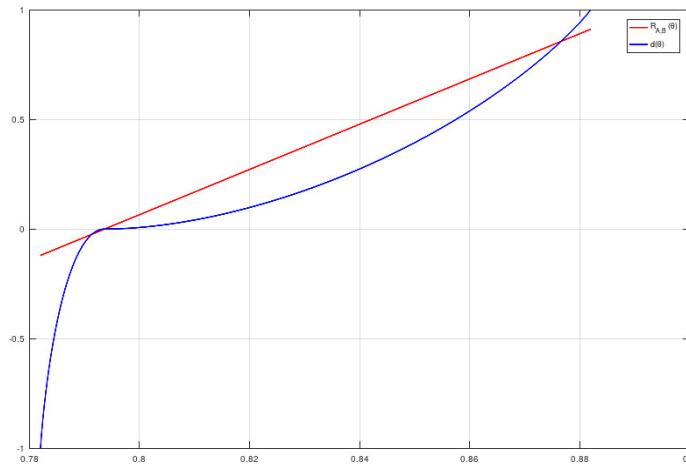
Figure 6

Comparison of the graphs of $e_{\alpha,\beta}(\theta)$ and $R_{\alpha;\beta}(\theta)$
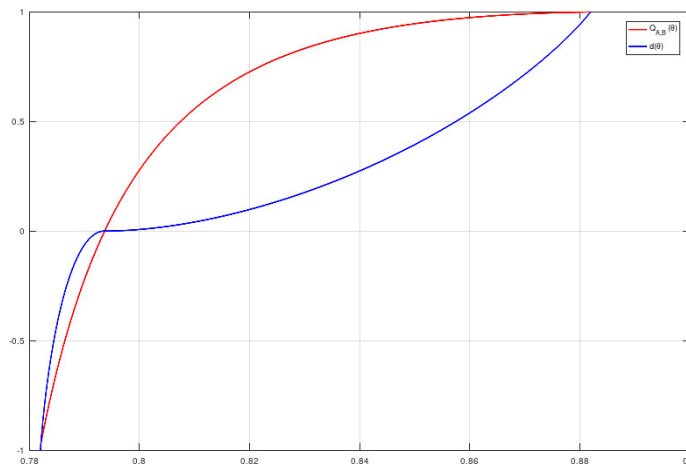
Figure 7

Comparison of the graphs of $e_{\alpha,\beta}(\theta)$ and $Q_{\alpha,\beta}(\theta)$

Acknowledgements

It is a pleasure for me to cordially thank Boyan Dimitrov, Kettering University, MI, USA, whose notes and suggestions are invaluable for making this

paper more readable.

I would like to thank both the administration of the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences and the administration of the American University in Bulgaria, for creating perfect and safe conditions of work.

This research received no external funding.

The author declares no conflict of interest.

References

[1] B. Dimitrov, Some Obreshkov Measures of Dependence and Their Use, Comptes rendus de l'Academie bulgare des Sciences 63, No 1 (2010), 5-18.

[2] I. M. Gelfand, A. N. Kolmogorov, A. M. Yaglom, Amount of Information and Entropy for Continuous Distributions, Mathematics and Its Applications, Selected Works of A. N. Kolmogorov, III: Information Theory and the Theory of Algorithms, pp. 33-56, Springer Science+Business Media Dordrecht 1993.

[3] A. N. Kolmogorov, Foundations of the Theory of Probability, Chelsea Publishing Company, New Yourk 1956.

[4] W. R. Macdonell, On the influence of previous vaccination in cases of small-pox, Biometrika, 1, No 3 (1902), 375-383.

[5] N. Obreshkoff, Theory of Probability, Nauka i Izkustvo, Sofia 1963 (In Bulgarian)

[6] C. E. Shannon, A Mathematical Theory of Communication, Bell System Technical Journal 27, No 3 (1948), 379-423.

[7] C. E. Shannon, A Mathematical Theory of Communication, Bell System Technical Journal 27, No 4 (1948), 623-656.

[8] G. Udny Yule, On the Association of the Attributes in Stattistics, Phil. Trans. Roy. Soc., A, 194, (1900), 257-319.

[9] G. Udny Yule, On the Methods of Measuring Association Between Two Attributes, Journal of the Royal Statistical Society, 75, No. 6 (1912), 579-652.