An Excursion in Finite-Field-Valued Measures

Erik I. Verriest[1] and Krishna Narayanan[1]

Received October 13, 1999

Probability and expectation are characterized within the framework of symmetric bilinear forms. Over the real field, the theory coincides with finite classical probability. In this paper we consider a theory over finite fields. For consistency, the probability measures must assume values in the finite field and obey the field operations. Thus, discrete probability theory is generalized in a new way. As a direct consequence, the sum of nonzero probabilities may become zero. Some of the implications, ramifications, problems, and results are explored.

1. INTRODUCTION: CLASSICAL PROBABILITY AND EXTENSIONS

Starting from some simple facts about classical finite probability, the ways in which classical probability can be generalized are indicated. One possible extension is singled out.

Consider a finite set $\Omega = \{1, 2, \ldots, n\}$; its elements are called *atoms*. A (classical) probability distribution P on a finite set is determined by the values on its atoms. Such a description will be called a *maximal* description. Given the probabilities $\{p_1, \ldots, p_n\}$, it also makes the set Ω *least redundant*. A meaningful interpretation requires that the probabilities p_i are nonnegative and sum to 1. The probability of an arbitrary subset of Ω is the sum of the probabilities of its atoms. This is equivalent to Kolmogorov's axiom

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B) \tag{1}$$

The *states* of (classical) physical systems are probability distributions. In a (classical) physical context, the *observables* are the random variables

[1] School of ECE, Georgia Institute of Technology, Atlanta, Georgia 30332-0250.

0020-7748/99/1200-3283$16.00/0 © 1999 Plenum Publishing Corporation

$$f: \quad \Omega \to \mathbb{R} \tag{2}$$

The *expectation* $\mathbf{E}f$ of a random variable f when the system is in state P is defined as

$$\mathbf{E}f = \Sigma p_j f(j) = [p_1, \ldots, p_n] \begin{bmatrix} f(1) \\ \vdots \\ f(n) \end{bmatrix} \tag{3}$$

This expresses the fact that the distribution belongs to the dual space $(\mathbb{R}^n)^*$. Alternatively the expectation may be expressed by

$$\mathbf{E}f = \mathrm{Tr} \begin{bmatrix} p_1 & & \\ & \ddots & \\ & & p_n \end{bmatrix} \begin{bmatrix} f(1) & & \\ & \ddots & \\ & & f(n) \end{bmatrix} \tag{4}$$

This formulates the distribution as an element of the dual of the n^2-dimensional space of Hermitian matrices. Finally, a third equivalent form is

$$\mathbf{E}f = [\sqrt{p_1}e^{-i\theta_1}, \ldots, \sqrt{p_n}e^{-i\theta_n}] \begin{bmatrix} f(1) & & \\ & \ddots & \\ & & f(n) \end{bmatrix} \begin{bmatrix} \sqrt{p_1}e^{i\theta_1} \\ \vdots \\ \sqrt{p_n}e^{i\theta_n} \end{bmatrix} \tag{5}$$

The extensions of classical probability have traditionally been via (4) and (5), by replacing the diagonal matrices by nondiagonal ones (Parthasarathy, 1992). See also Gudder (1988).

The novel viewpoint explored in this paper starts from an extension based on (3). As further motivation, a notion of correlation between observables is taken as a point of departure. The *correlation* between two observables f and g is defined as

$$C_{f,g} = \langle f, g \rangle = [f(1), \ldots, f(n)] \begin{bmatrix} p_1 & & \\ & \ddots & \\ & & p_n \end{bmatrix} \begin{bmatrix} g(1) \\ \vdots \\ g(n) \end{bmatrix} \tag{6}$$

By introducing the *unit function* $\mathbb{1}$

$$\mathbb{1}: \Omega \to \mathbb{R}: \mathbb{1}(\omega) = 1, \quad \forall \omega \in \Omega \tag{7}$$

we recover the expectation as

$$\mathbf{E}f = \langle f, \mathbb{1} \rangle \tag{8}$$

The correlation in the classical theory, (6), is an inner product in $\mathbb{R}^n$, the space of $\mathbb{R}$-valued random variables over Ω, with $|\Omega| = n$.

Guided by this motivation, we seek to generate a theory over arbitrary finite fields. In the next section, we set up this framework more precisely.

2. SYMMETRIC BILINEAR FORM

Let $\mathbb{F}$ be a finite field, and V a vector space over $\mathbb{F}$. The addition and multiplication in $\mathbb{F}$ will be denoted by $\oplus$ and $\otimes$, respectively.

Taking correlations as a point of departure, one would like to define an *inner product* in V as a map $(x, y) \rightarrow \langle x, y \rangle$ from $V \times V$ to a number field. Guided by classical theory, desirable properties for such an inner product are as follows:

(i) Linearity in its second argument.
(ii) $\langle x, y \rangle = \langle y, x \rangle$ (noncomplex).
(iii) $\langle x, x \rangle \geq 0$, with equality iff $x = 0$.

If we let the bilinear form take values in $\mathbb{R}$, then linearity implies for $x, y, z \in \mathbb{F}$

$$\langle x \oplus y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \tag{9}$$

which for $\mathbb{F} = \mathbf{GF}(3)$ yields

$$\langle 2 \oplus 1, z \rangle = \langle 2, z \rangle + \langle 1, z \rangle$$

But the left-hand side is $\langle 0, z \rangle = 0$, whereas the right-hand side is $2z + z = 3z$, which is inconsistent. Hence for linearity to hold, the image number field must be the base field of the vector space, i.e., $\mathbb{F}$.

Now since there is no consistent order relation in a finite field $\mathbb{F}$, one cannot speak about "positivity." Therefore, the third desirable property needs to be dropped. Can the inequality be relaxed to nonequality? This also fails. In the standard representation of a vector x with respect to a basis as a column of scalars $[x_1, x_2, x_3, \ldots]'$, where the prime denotes the transpose that converts columns to rows and vice versa, then $\langle x, y \rangle = \Sigma_i\, x_i \otimes y_i$ is an inner product. The vectors x in the linear span of $[1, 1, 1]' \in \mathbf{GF}(3)^3$ all satisfy $\langle x, x \rangle = 0$, whereas both $[1, 1, 1]$ and $[2, 2, 2]$ are obviously nonzero. A space with a symmetric bilinear form shares some properties of a Krein space (Bognár, 1974; Gohberg *et al.*, 1983), which is a vector space over $\mathbb{R}$ or $\mathbb{C}$, on which an *indefinite* inner product is defined. However, notice that in the finite field context, the question of definiteness itself is meaningless.

Thus we define a symmetric bilinear form on V as a map $(x, y) \rightarrow \langle x, y \rangle$ of $V \times V$ into $\mathbb{F}$ such that for any $y \in V$ the map

$$y_R\colon \quad x \rightarrow \langle x,y \rangle$$

is a linear function from V to $\mathbb{F}$, and likewise, for any $x \in V$, the map

$$x_L: \quad y \to \langle x, y \rangle$$

is a linear function from V to $\mathbb{F}$. The geometric properties of such spaces are described by O'Meara (1963) and Jacobson (1985).

Now why may finite fields be of interest? Any system that can be descibed by an *automaton* has a local representation over $\mathbb{F}$. Applications of deterministic linear automata abound in coding, cryptography and symbolic dynamics (Booth, 1967; Gill, 1962; Pretzel, 1996).

Figure 1 shows a "black box" representation of a deterministic automaton having a finite number of possible states. At time k, let x_k be the state of the automaton, described by a vector in $\mathbb{F}^n$. The input (or control or conditioning) u_k changes the state of the automaton according to the $\mathbb{F}$-linear dynamical equation

$$x_{k+1} = Ax_k \oplus bu_k \tag{1 0}$$

where $A \in \mathbb{F}^{n \times n}$ and $b \in \mathbb{F}^{n \times 1}$. Let us also assume that the output (or measurement) of the automaton at time k depends on the state and the input through

$$y_k = cx_k \oplus du_k \tag{11}$$

where $c \in \mathbb{F}^{n \times 1}$ and $d \in \mathbb{F}$. Extensions to multiinput, multioutput automata are straightforward (Booth, 1967). The initial state x_0 and the sequence of inputs $\{u_k\}$ produces thus the evolution of the system and determines (uniquely) its output sequence $\{y_k\}$.

For a stochastic automaton, a probabilistic description of the above is given. The initial state and inputs are random variables. Assuming statistical independence between the initial state and the successive inputs, a *big* product space $\Omega_{-1} \times \Omega_0 \times \Omega_1 \times \ldots$ is introduced, where $x_0 \in \Omega_{-1}$, $u_0 \in \Omega_0$, $u_1 \in \Omega_1, \ldots$ Since all intervening random variables in such a state space model are finite-field-valued, a description defining correlations by

$$C_{f,g} = \langle f, g \rangle = [f(1), \ldots, f(n)] \begin{bmatrix} p_1 & & \\ & \ddots & \\ & & p_n \end{bmatrix} \begin{bmatrix} g(1) \\ \vdots \\ g(n) \end{bmatrix} \tag{12}$$

must be consistent in $\mathbb{F}$. Hence the probabilities themselves must be $\mathbb{F}$-valued.

The normalization condition follows from $\mathbf{E}\mathbb{1} = 1$, i.e.,

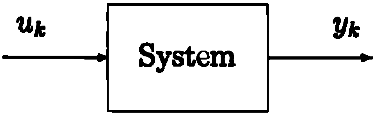


Fig. 1. Automaton.

$$\mathbf{E}\mathbb{1} = \mathbb{1}'P\mathbb{1} = \operatorname{Tr} P = 1 \tag{13}$$

Thus

$$p_1 \oplus p_2 \oplus \ldots \oplus p_n = 1 \tag{14}$$

Example. A scalar system in **GF**(3):

$$x_{k+1} = x_k \oplus u_k$$

$$y_k = x_k$$

With $\Omega_{u_k} = \{0, 1, 2\}$, there are $3^3 = 27$ different random variables (observables). The three deterministic (or pure) states $x_0 \in \{0, 1, 2\}$ are imbedded as

$$x_0 \rightarrow \{x | x(\omega) = x_0, \forall \omega\}$$

Likewise, there are $3^2 = 9$ possible states (measures) on Ω_{u_k}. These are:

1. The three permutations of $[1, 0, 0]'$, which are the "pure states."
2. The three permutations of $[2, 2, 0]'$.
3. The three permutations of $[1, 1, 2]'$.

Introducing the equivalence relation $\sim$ among $\mathbb{F}$-valued random variables,

$$x \sim y \Leftrightarrow P(\{\omega | x(\omega) \neq y(\omega)\}) = 0 \tag{15}$$

it follows that:

(i) If P is of the first form, say $P = \operatorname{diag}\{1, 0, 0\}$, there are three equivalence classes. We group them as class I:

$$[i] := \left\{ \begin{bmatrix} i \\ \times \\ \times \end{bmatrix}, \quad i = 0, 1, 2 \right\} \tag{16}$$

(ii) With the probability measures of type $P = \operatorname{diag}\{2, 2, 0\}$, there correspond six equivalence classes, characterized by a two-dimensional set, grouped as class II:

$$\begin{bmatrix} i \\ j \end{bmatrix} := \left\{ \begin{bmatrix} i \\ j \\ \times \end{bmatrix}, \quad i, j \in \{0, 1, 2\} \right\} \tag{17}$$

(iii) The measures of type $P = \operatorname{diag}\{1, 1, 2\}$ determine a three-dimensional set, class III. For instance, the different classes of *zero-mean* variables in **GF**(3) are characterized as

Class I: $[0]$

Class II: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Class III: $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, & permutations

For instance, with probability $P = \text{diag}\{2, 2, 0\}$ in $\mathbf{GF}(3)$, there are four subspaces of random variables. Denoting the $\mathbb{F}$-linear span by $\mathcal{L}$, these are

$$\mathcal{L}\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathcal{L}\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathcal{L}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathcal{L}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Defining $x \perp y$ by the condition $C_{x,y} = 0$, we get

$$\mathcal{L}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \perp \mathcal{L}\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathcal{L}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \perp \mathcal{L}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

3. THE AXIOMATIC SETUP

As in classical probability theory, we take for the space of *outcomes* (the sample space) a *finite* set $\Omega = \{1, 2, \ldots, n\}$. The *events* in the theory are the elements of a (Boolean) algebra $\mathcal{B}$ of subsets of Ω closed under union and complementation and containing Ω. The underlying *logic* in this theory is therefore Boolean. Since Ω is finite, there is no need to introduce a σ-algebra to define a *measurable space* $(\Omega, \mathcal{B})$. The subset $[\mathcal{B}] = \{B_i \in \Omega\}$ of the algebra $\mathcal{B}$ is called a *generating partition* for $\mathcal{B}$ if $\cup B_i = \Omega$ and for $i \neq j$, $B_i \cap B_j = \emptyset$. Notice that the atoms of Ω form a generating partition iff $\mathcal{B} = 2^{\Omega}$, the power set of Ω. Under this condition, $(\Omega, \mathcal{B})$ is least redundant. The deviation from the classical theory comes from the measures.

Definition 3.1. Given a finite field $\mathbb{F}$, a Galois Measure Q on the measurable space $(\Omega, \mathcal{B})$ is an $\mathbb{F}$-valued additive mapping Q: $(\Omega, \mathcal{B}) \to \mathbb{F}$ satisfying:

(i) $Q(\Omega) = 1$.
(ii) If $A, B \in \mathcal{B}$ with $A \cap B = \emptyset$, then $Q(A \cup B) = Q(A) \oplus Q(B)$.

Note that the first axiom is just a normalization condition. These two axioms are the same as required in the Kolmogorov definition of probability. The axiom of *positivity* in the Kolmogorov theory is nonexistent here since the Galois field is not an ordered field. If $\mathbb{F} = \mathbf{GF}(p)$ for p prime, we shall denote the measure by Q_p. By analogy with the word "probability," the term *galoibility* (after Galois) will also be used to denote the Galois measure of an event. Note that the Galois measure of the measurable sets in $\mathcal{B}$ is completely determined by the galoibility on the generating partition.

It follows from definition (3.1) that

$$Q(\emptyset) = 0$$

Denoting by $\overline{A}$ the complement of A in Ω, then for $Q = Q_p$, we have

$$Q_p(\overline{A}) = 1 \oplus (p - 1)Q_p(A)$$

since $p - 1$ is the additive inverse of 1 in $\mathbf{GF}(p)$. In general, one has

$$Q_p(A \cup B) = Q_p(A) \oplus Q_p(B) + (p - 1)Q_p(A \cap B)$$

An important consequence of this is that only sets with $kp + 1$ elements, $k \geq 0$ integer, can have a 'uniform' distribution Q_p. In particular if $\Omega = \mathbf{GF}(p)$ is used for instance as trivial sample space of a $\mathbf{GF}(p)$-valued random variable, a 'uniform' distribution is nonexistent since it cannot be normalized: "*Galois probability abhors uniformity.*"

Theorem 3.2. There are p^{p-1} different Galois measures on $\Omega = \mathbf{GF}(p)$ with its maximal algebra.

Proof. Follows from a simple counting argument.

Definition 3.3. A measurable space endowed with a Galois measure will be called a Galois measure space, and denoted by $(\Omega, \mathcal{B}, Q)$.

Definition 3.4. Let $(\Omega, \mathcal{B}, Q)$ be a Galois measure space; the event $B \in \mathcal{B}$ is called *certain* if $Q(B) = 1$ and no subset of its complement B has nonzero galoibility. Event $B \in \mathcal{B}$ is *quasicertain* if $Q(B) = 1$ and $\exists A \subset \overline{B}$ such that $Q(A) \neq 0$. If $Q(B) \neq 1$, the event B is called (galois) *probable*.

Example. Let $\Omega = \{0, 1, 2, 3, 4\}$, with Q represented by diag$\{1, 1, 1, 3, 0\}$ in $\mathbf{GF}(5)$, the events $\{0\}$, $\{1\}$, and $\{2\}$ are quasicertain, the event $\{0, 1, 2, 3\}$ is certain, and $\{3\}$ is probable. Notice that a certain event may itself be composed of quasicertain subsets.

Definition 3.5. A Galois random variable (grv) x on a Galois measure space $(\Omega, \mathcal{B}, Q)$ is an $\mathbb{F}$-valued, $\mathcal{B}$-measurable map x: $\Omega \to \mathbb{F}$, i.e.,

$$x^{-1}(f) \subset \mathcal{B}, \quad \forall f \in \mathbb{F}$$

Then x is said to assume the value f with galoibility $Q(x^{-1}(f))$. Two random variables, x and y on a Galois measure space are *equivalent* if $Q\{\omega | x(\omega) \neq y(\omega)\} = 0$.

If $|\Omega| = n$, the grv's are represented as vectors in the space $\mathbb{F}^n$, endowed with a symmetric bilinear form

$$\langle\cdot,\cdot\rangle\colon \mathbb{F}^n \times \mathbb{F}^n \to \mathbf{GF}(p)\colon (x, y) \to \sum_{i=1}^{n} x_i \otimes y_i \otimes Q(i) = \langle x, y\rangle \qquad (18)$$

representing the correlation $C_{x,y}$.

The *expectation* of x is defined by the "integral" with respect to the Galois measure, and is obtained by $C_{x,1}$. Thus the Galois random variable x has Q_p-expectation

$$\int_{\Omega} x(\omega) Q_p(\omega) = \sum_{i=0}^{n} x_i \otimes Q_p(i)$$

4. CONDITIONING

Let $B \in \mathcal{B}$, with $Q(B) \neq 0$. The classical definition of conditional probability can be carried over directly.

Definition 4.1. A conditional Galois measure $Q(\cdot|B)$ on $(\Omega, \mathcal{B})$ is a Galois measure on $(\Omega, \mathcal{B})$ satisfying

$$Q(A|B) = \frac{Q(A \cap B)}{Q(B)}, \qquad \forall A \subset \mathcal{B} \qquad (19)$$

In terms of the indicator function and correlations the definition reduces to

$$Q(A|B) = \frac{\sum \chi_A(i)\chi_B(i)Q(i)}{\sum \chi_B(i)Q(i)} = \frac{C_{\chi_A,\chi_B}}{C_{\chi_B,1}} \qquad (20)$$

For instance, with $(\Omega, \mathcal{B}, Q) = (\mathbf{GF}(3), 2^{\mathbf{GF}(3)}, \text{diag}\{2, 1, 1\})$ one finds $Q(\cdot|\{1, 2\}) = \text{diag}\{0, 2, 2\}$.

Conditional expectation is expectation with respect to the conditional measure. Consider the Galois measure space in the above example. For the grv $x = [x_0, x_1, x_2]'$, one finds $\mathbf{E}(x) = 2x_0 \oplus x_1 \oplus x_2$ and $\mathbf{E}(x|\{1, 2\}) = 2x_1 \oplus 2x_2$, $\mathbf{E}(x|\{2\}) = x_2$, $\mathbf{E}(x|\{0\}) = x_0$. The quantity $\mathbf{E}(x|\{0, 1\})$ cannot be defined via conditional probability since $Q(\{0, 1\}) = 0$. Can the notion of conditional expectation be defined as Kolmogorov did, thus making conditional expectation more fundamental than conditional probability? To this end, consider all measurable functions $(\Omega, \mathcal{B}) \to \mathbb{F}$ for an arbitrary algebra $\mathcal{B}$.

Definition 4.2. The *conditional expectation* of a grv x with respect to a subalgebra $\mathcal{G}$ of $\mathcal{B}$ is the $\mathcal{G}$-measurable grv, denoted $\mathbf{E}(x|\mathcal{G})$ or $\mathbf{E}^{\mathcal{G}}x$, such that for every $G \in \mathcal{B}$, and every $\mathcal{G}$-measurable grv ϕ one has

$$\int_G x\phi \, dQ = \int_G \mathbf{E}(x|\mathcal{G})\phi \, dQ \tag{21}$$

Note that this defines a unique conditional expectation, if it exists, under the equivalence $\sim$. This definition can be relaxed, since in the finite case all measurable functions are generated by the indicator functions. Thus, using the correlations,

$$\forall G \in \mathcal{G}: \quad C_{x,\chi_G} = C_{E^{\mathcal{G}}x,\chi_G} \tag{22}$$

In the above example, let $B = \{0, 1\}$ and $\mathcal{G} = \{\emptyset, B, \overline{B}, \mathbf{GF}(3)\}$. The indicator functions of interest are the functions taking values (1, 0) and (0, 1), respectively, for (B, B). If the conditional expectation assumes values X on B and X on B, then X and X need to satisfy the relations

$$2x_0 \oplus x_1 = X \otimes (2 \oplus 1)$$
$$x_2 = \overline{X}$$

Notice that the first condition implies that, unless $2x_0 + x_1 = 0$, there cannot be a consistent definition of the conditional expectation following Kolmogorov. More generally, we have the following:

Theorem 4.3. Let $(\Omega, \mathcal{B}, Q_p)$ be a Galois measure space, and let x be a grv defined on this space. Given a subalgebra $\mathcal{G}$ of $\mathcal{B}$, the conditional expectation $\mathbf{E}^{\mathcal{G}}x$ exists if for every G in the generating partition $[\mathcal{G}]$ such that $Q(G) = 0$, the correlation $C_{x,\chi_G} = 0$. In this case the conditional expectation on probable G's is given by

$$E(x|G) = \begin{cases} \dfrac{C_{x,\chi_G}}{Q(G)} & \text{if} \quad Q(G) \neq 0 \\ \text{arbitary} & \text{if} \quad Q(G) = 0 \end{cases}$$

Proof. Generalizing the example, we have for the sets in the generating partition $[\mathcal{G}]$,

$$C_{x,\chi_G} = C_{\mathbf{E}^{\mathcal{G}},G} = E(x|G)Q(G)$$

from which the statement follows. ∎

If B is certain and $\mathcal{G}$ is the algebra generated by B and any subalgebra of its complement, then $\mathbf{E}^{\mathcal{G}}x = \mathbf{E}x$.

5. BERNOULLI TRIALS

The description of repeated experiments, all physically independent, requires the definition of product spaces. Given a Galois measure space $(\Omega,$

$\mathcal{B}$, Q), consider the Cartesian product $\Omega\times\Omega\times \cdots \times\Omega$ (N copies), together with the algebra generated by $\mathcal{B}\times\mathcal{B}\times\cdots\times\mathcal{B}$. Denote them respectively by Ω^N and $\mathcal{B}^N$. Define a mapping on the product measurable space (Ω^N, $\mathcal{B}^N$) into $\mathbb{F}$ as follows:

$$\hat{Q}(B_1 \times B_2 \times \cdots \times B_N) = Q(B_1) \otimes Q(B_2) \otimes \cdots \otimes Q(B_N) \qquad (23)$$

Then $\hat{Q}$ satisfies

(i) The normalization condition $\hat{Q}$ (Ω^N) $= Q(\Omega)^{\otimes N} = 1$.

(ii) On the generating algebra of (Ω^2, $\mathcal{B}^2$), if $B_1 \cap B_2 = B_1' \cap B_2' = \emptyset$, then

$$(B_1 \times B_2) \cup (B_1' \times B_2') = (B_1 \cup B_1') \times (B_2 \cup B_2')\backslash\ (B_1' \times B_2)\backslash(B_1 \times B_2')$$

and therefore it follows from the galoibility axioms that

$$\hat{Q}((B_1 \times B_2) \cup (B_1' \times B_2')) = \hat{Q}(B_1 \times B_2) \oplus \hat{Q}(B_1' \times B_2')$$

The $\mathbb{F}$-additivity in the general case for (Ω^N, $\mathcal{B}^N$) easily follows by induction. Hence $\hat{Q}$ is a valid Galois measure making (Ω^N, $\mathcal{B}^N$, $\hat{Q}$) into a Cartesian product Galois measure space. $\hat{Q}$ is a product measure induced by Q. In the same fashion, the product of different Galois measure spaces can be formed to define a galoibility space associated with general combined physically independent experiments. By a *Bernoulli experiment*, we refer to independent trials of a yes/no experiment, that is, one with $\mathcal{B} = \{\emptyset, B, B, \Omega\}$. As defined, all random variables described over such a product space must be $\mathbb{F}$-valued. Now one "obvious" random variable one may want to consider is the *number of successes* in N successive trials. However, such a count necessarily assumes all possible integer values from 0 to N. There is no room for such a thing in this framework. In fact, at a more fundamental level, we also have no room to even consider N repetitions for $N \geq p$ in $\mathbb{F} = \mathbf{GF}(p)$, since in this universe, one simply cannot count higher than $p - 1$. This is temporarily sidestepped by allowing a natural number-valued count, referred to as a *count*. (It will be shown later that this is indeed unnecessary.) The definition of *count*, which falls outside the domain of $\mathbb{F}$, is only justifiable using the product space. For instance in $\mathbf{GF}(3)$, the event {count $= 4$} is the event

$$(B \times B \times B \times \overline{B}) \cup (B \times B \times \overline{B} \times B) \cup (B \times \overline{B} \times B \times B)$$
$$\cup\ (\overline{B} \times B \times B \times B)$$

and is thus well defined. Within the finite-field scope, only the *intracount* can be defined. By this we shall understand the $\mathbb{F}$-valued count $\hat{k}(\omega)$ [i.e., "count" modulo p in $\mathbf{GF}(p)$]. Only the intracount can be measured and is a

$(\Omega^N, \mathcal{B}^N, \hat{Q}_N)$-grv. If $k_N(\omega)$ is the natural-valued count given N, then $\hat{k}_N(\omega) = k_N(\omega) \bmod p$ is the corresponding intracount. Clearly,

$$\hat{k}_N = f \Leftrightarrow \text{outcome } (\omega_1, \omega_2, \ldots, \omega_N) \text{ contains } f, f + p, f + 2p, \ldots, f + lp \text{ components in } B \tag{24}$$

where $f + lp \leq N < f + (l + 1)p$, or l is the largest integer smaller than N/p.

Theorem 5.1. Given a **GF**(p)-Bernoulli trial with $Q(success) = a$, then the intracount distribution is p-periodic in N, the number of trials. Moreover if $N_0 < p$, then the galoibility of the intracount in the Bernoulli experiment is given by

$$\hat{Q}_{N_0}(\hat{k}) = \binom{N_0}{\hat{k}} a^{\hat{k}} (1 - a)^{N_0 - \hat{k}} \bmod p \tag{25}$$

The event "K successes in N trials" has galoibility

$$\hat{Q}_N([K]_N) = \prod_{i=0}^{\alpha} \hat{Q}_{N_i}(k_i)$$

expressed in terms of the p-ary expansion of K and N given by

$$N = N_0 + N_1 p + N_2 p^2 + \cdots + N_\alpha p^\alpha \tag{26}$$

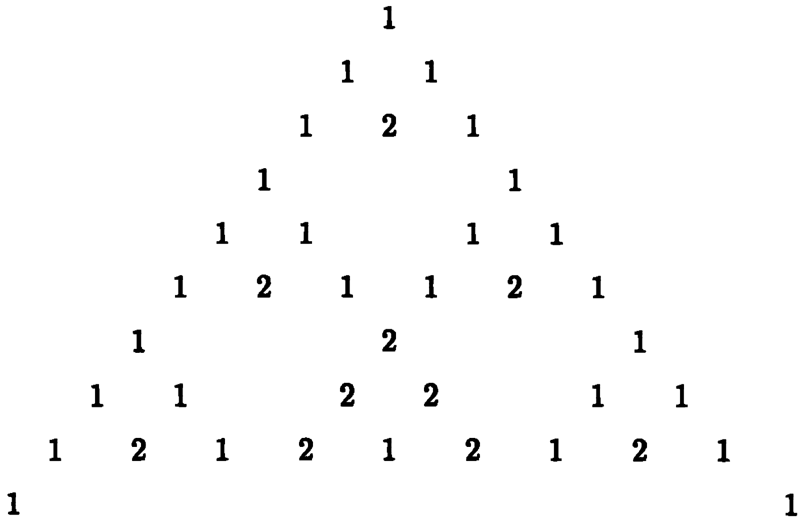
$$k = k_0 + k_1 p + k_2 p^2 + \cdots + k_\alpha p^\alpha \tag{27}$$

Proof. See Appendix.

A direct consequence of this theorem is that indeed in this space there is no reason to count higher than $p - 1$, as mentioned before. The count does not provide any information that is not already present in the intracount.

There is a simple visualization of this result: Write down Pascal's triangle modulo p. A self-similar structure, the Sierpinski triangle, results. The self-similarity arises from there petition of the factors $\binom{N_i}{k_i}$ as N is increased. The Sierpinski triangle is shown in Fig. 2, where for clarity the 0's are omitted.

As a final observation, note that the expectation of $\hat{k}_N$ evaluates, as expected, to $aN \bmod p$. There is, however, no relative frequency notion which gives a simple interpretation of this. For instance, in **GF**(5), one finds for $a = 2$, $\mathbf{E}\hat{k}_1 = 2$ and $\mathbf{E}\hat{k}_2 = 4$. There can be no "limit theorems" in a theory over finite fields.



```
                  1
                1   1
              1   2   1
            1           1
          1   1       1   1
        1   2   1   1   2   1
      1           2           1
    1   1       2   2       1   1
  1   2   1   2   1   2   1   2   1
1                                   1
```

Fig. 2. Pascal triangle in **GF**(3).

6. RANDOM WALK

In order to illustrate the idea, consider a random walk on **GF**(3), starting at 0. Let the individual steps u_k be independent and identically distributed (i.i.d) with zero mean. This implies that for all k, $Q(u_k = 1) = Q(u_k = 2) = 2$, since $2 \oplus 2 = 1$ in **GF**(3).

We get, with $x_0 = 0$, a pure state, the successive galoibilities of the position x_k:

$$x_0 = 0 \Rightarrow \begin{bmatrix} x_k(2) \\ x_k(0) \\ x_k(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$x_1 = u_1 \Rightarrow \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$$

$$x_2 = u_1 \oplus u_2 \Rightarrow \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$x_3 = u_1 \oplus u_2 \oplus u_3 \Rightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

We departed from the "ordering" (0, 1, 2), in order to keep the mean in the "middle" of Fig. 3. [Recall that $2 = -1$ in **GF**(3).] The numbers above the ○ are the galoibilities.

As shown in Fig. 3, the process is self-focusing or repetitive. An alternative interpretation is that *time itself is recycled.* This curious behavior is not an artifact of **GF**(3), but is true for all **GF**(p).

Theorem 6.1. For $p \geq 2$, the random walk in **GF**(p) is recycled (i.e., refocuses into its initial state after p transitions.

Proof. Since p must be odd, set $p = 2q + 1$. Then, using -1 to denote the field element $p - 1$, the random walk process is characterized by

$$Q(u = 1) = Q(u = -1) = \frac{1}{2} \bmod p = q + 1 \tag{28}$$

The Markov chain transition matrix for the state of the random walk is then

$$M = (q + 1)\begin{bmatrix} 0 & 1 & & & & 1 \\ 1 & \ddots & \ddots & & \bigcirc & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & \bigcirc & & \ddots & \ddots & 1 \\ 1 & & & & 1 & 0 \end{bmatrix} \tag{29}$$

The k-step transition probability is M^k, and thus the probability distribution at time k is $\pi_k = M^k \pi_0$. Now, by the Cayley–Hamilton theorem, $M^p = \mathrm{I}$, thus proving the assertion. ■

It follows from the above theorem also that stationary Galois distributions do not exist. The one-dimensional random walk has p^{p-1} possible states. Each state belongs to a p-cycle, thus creating p^{p-2} cycles in the state transition

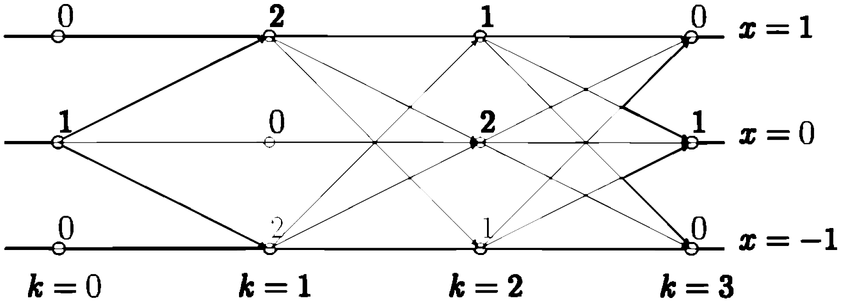


Fig. 3. Random walk in **GF**(3).

diagram for a random walk in **GF**(p). Modulo cyclic permutations, this leaves p^{p-3} pattern cycles.

7. CONCLUSIONS

Some initial ideas were presented toward a generalization of probability theory by letting probabilities assume values in a finite field. It was discovered that an elementary symmetric random walk in **GF**(p) refocuses after p steps. If $p = 2q + 1$ is large, then for $k \leq q$ steps, the random walk is indistinguishable from the random walk described on an integer lattice. Because of the repetitive behavior, convergence theorems for galoibilities similar to classical probability (i.e., law of large numbers) are nonexistent.

APPENDIX

Lemma A.1 Let $\overset{p}{\equiv}$ denote congruence modulo p. Let also $\lfloor\cdot\rfloor$ denote the floor-function: $\lfloor x\rfloor$ is the largest integer smaller than or equal to x. For any two positive integers a and b:

$$\frac{(a+1)(a+2)\cdots(a+p)}{(b+1)(b+2)\cdots(b+p)} \overset{p}{\equiv} \frac{\lfloor a/p\rfloor + 1}{\lfloor b/p\rfloor + 1}$$

Proof. Any set of p consecutive integers contains a multiple of p. Using the division theorem, $a = \alpha_1 p + \alpha_0$, $b = \beta_1 p + \beta_0$ with $0 \leq \alpha_0, \beta_0 < p$, and canceling the p between numerator and denominator gives

$$\begin{aligned}\frac{(a+1)(a+2)\cdots(a+p)}{(b+1)(b+2)\cdots(b+p)} &= \frac{(\alpha_1 p + \alpha_0 + 1)(\alpha_1 p + \alpha_0 + 2)\cdots((\alpha_1 + 1)p)\cdots((\alpha_1+1)p + \alpha_0)}{(\beta_1 p + \beta_0 + 1)(\beta_1 p + \beta_0 + 2)\cdots((\beta_1 + 1)p)\cdots((\beta_1+1)p + \beta_0)} \\ &\overset{p}{\equiv} \frac{\alpha_1 + 1}{\beta_1 + 1}\frac{(p-1)!}{(p-1)!}\end{aligned}$$

By Wilson's theorem (Barnett, 1972), $(p-1)! \overset{p}{\equiv} p - 1$, the above fraction is well defined, thus proving the lemma. ■

Simplify the notation by defining $\phi_x = \lfloor x/p\rfloor$. We have the following result.

Lemma A.2. For positive integers N, n, and k, there holds

$$\binom{N+np}{k} \overset{p}{\equiv} \frac{(\phi_N+n)!\phi_{N-k}!}{(\phi_{N-k}+n)!\phi_N!}$$

Proof. Consider first

$$\binom{k+m+p}{k} = \frac{(k+m+p)\cdots(k+m+1)}{(m+p)\cdots(m+1)}\binom{k+m}{k}$$
$$= \frac{\phi_{k+m}+1}{\phi_m+1}\binom{k+m}{k}$$

by Lemma A.1. Hence one deduces the recursion

$$\binom{k+m+np}{k} = \frac{\phi_{k+m}+n}{\phi_m+n}\binom{k+m+(n-1)p}{k}$$

which upon iterating gives

$$\binom{k+m+np}{k} = \frac{\phi_{k+m}+n}{\phi_m+n}\,\frac{\phi_{k+m}+n-1}{\phi_m+n-1}\cdots\frac{\phi_{k+m}+1}{\phi_m+1}\binom{k+m}{k}$$

from which the statement follows. ■

Lemma A.3. If N, n, k, and l are nonnegative integers, with N and k less than p, then

$$\binom{N+np}{k+lp} \overset{p}{\equiv} \binom{n}{l}\binom{N}{k}$$

Proof. Note that

$$\binom{k+m+lp}{k+lp} = \binom{k+m+lp}{m} \overset{p}{\equiv} \frac{(\phi_{k+m}+l)!\phi_k!}{(\phi_k+l)!\ \phi_{k+m}!}\binom{k+m}{m} \tag{30}$$

by Lemma A.2. It follows then from Lemma A.2 and (30) that for $n \geq l$

$$\binom{N+np}{k+lp} = \binom{N+(n-l)p+lp}{k+lp}$$
$$\overset{p}{\equiv} \frac{(\phi_{N+(n-l)p}+l)!\phi_k!}{(\phi_k+l)!\ (\phi_{N+(n-l)p})!}\binom{N+(n-l)p}{k}$$
$$\overset{p}{\equiv} \frac{(\phi_N+n)!\phi_k!}{(\phi_k+l)!(\phi_N+n-l)!}\binom{N+(n-l)p}{k}$$

If $N \geq k$, this reduces further to

$$\overset{p}{\equiv} \frac{(\phi_N + n)!\phi_k!}{(\phi_k + l)!(\phi_N + n - l)!} \frac{(\phi_N + n - l)!\phi_{N-k}!}{\phi_N!(\phi_{N-k} + n - l)!} \binom{N}{k}$$

$$\overset{p}{\equiv} \frac{\binom{\phi_N + n}{l}}{\binom{\phi_k + l}{l}} \frac{\binom{\phi_k + n - l}{n - l}}{\binom{\phi_{N-k} + n - l}{n - l}} \binom{N}{k}$$

If now $0 \leq k \leq N < p$, then $\phi_N = \phi_k = \phi_{N-k} = 0$, and thus

$$\binom{N + np}{k + lp} \overset{p}{\equiv} \binom{n}{l}\binom{N}{k}$$

In the other cases ($N < k$ with $n \geq l$ and $n < l$) it is readily verified that

$$\binom{N + np}{k + lp} \overset{p}{\equiv} 0$$

Hence in all cases, the stated formula has been shown. ■

Lemma A.4. Let N and k be positive integers with p-ary expansions (26) and (27), with at least one of k_α and N_α nonzero; then

$$\binom{N}{k} \overset{p}{\equiv} \prod_{i=0}^{\alpha} \binom{N_i}{k_i} \tag{31}$$

Proof. This is a direct consequence of Lemma A.3. ■

Note that if one of the N_i is smaller than the corresponding k_i, the binomial $\binom{N}{k}$ is congruent (modulo p) to zero.

Lemma A.5. Let N and k be positive integers with p-ary expansions given in Lemma A.4). Then for any integer $0 < a < p$, the following congruence holds:

$$a^k(1 - a)^{N-k} \overset{p}{\equiv} \prod_{i=0}^{\alpha} a^{k_i}(1 - a)^{N_i - k_i} \tag{32}$$

Proof. By Fermat's (little) theorem (Barnett, 1972), if $k = lp + k_0$ and $N = \nu p + N_0$, with k_0, $N_0 < p$, then

$$\begin{aligned} a^k(1 - a)^{N-k} &= a^{lp+k_0}\,(1 - a)^{(\nu-l)p+N_0-k_0} \bmod p \\ &= (a^l)^p a^{k_0}((1 - a)^{\nu-l})^p(1 - a)^{N_0-k_0} \bmod p \end{aligned}$$

$$= a^l(1 - a)^{\nu - l} a^{k_0}(1 - a)^{N_0 - k_0} \bmod p$$

By induction, the assertion follows. ■

Proof of Theorem 5.1. By the division algorithm, $N = N_0 + np$ with $N_0 < p$. Represent the event "count $= K$" or K successes in the N trials for notational simplicity as $[K]_N$. Recall that the event $[K]_N$ is well defined, but K is not a grv. Let also $K = k + ip$, with $k \leq p$.

It follows from Lemmas A.3 and A.5 that

$$\hat{Q}_N([k + ip]_N) \overset{p}{\equiv} \binom{N_0 + np}{k + ip} a^{k+ip}(1 - a)^{N_0 - k + (n-i)p}$$

$$\overset{p}{\equiv} \left[\binom{N_0}{k} a^k (1 - a)^{N_0 - k}\right]\left[\binom{n}{i} a^i (1 - a)^{n-i}\right]$$

The iterative form in Lemmas A.3 and A.4 yields the formula for the galoibility associated with event $[K]_N$, by noting that for all N_i, $k_i < p$, count and intracount are identical and $\hat{Q}_{N_i}([k_i]) = \hat{Q}_{N_i}(k_i)$.

Now consider the event "intracount in N trials is k," $k < p$, denoted by $\hat{k}_N$. Recall that this defines a grv and that

$$\hat{k}_N = [k]_N \cup [k + p]_N \cup \cdots \cup [k + l_p]_N$$

with l depending on N, such that $N - p < k + l_N p \leq N$. Then

$$\hat{Q}_N(\hat{k}_N) = \sum_{i=0}^{l} \hat{Q}_N([k + ip]_N)$$

$$= \hat{Q}_{N_0}(k) \sum_{i=0}^{l} \binom{n}{i} a^i (1 - a)^{n-i}$$

It is readily seen that if $k \leq N_0$, then $l = n$, and by the binomial theorem, the summation is congruent to 1, thus leading to

$$\hat{Q}_N(\hat{k}_N) = \hat{Q}_{N_0}(k) \tag{33}$$

If $N_0 < k$, then $l = n - 1$, and the summation evaluates to $1 - a^n$. However, in this case the factor $\hat{Q}_{N_0}(k)$ is zero because $\binom{N_0}{k} = 0$. Thus the equality (33) is universally valid. It shows that the distribution of $\hat{k}_N$ only depends on N through the residual N_0, thus proving the periodicity. ■

REFERENCES

Barnett, I. A., *Elements of Number Theory*, Prindle, Weber and Schmidt, 1972.

Bognár, J., *Indefinite Inner Product Spaces*, Springer-Verlag, 1974.

Booth, T., *Sequential Machines and Automata Theory*, Wiley, 1967.

Gill, A., *Introduction to the Theory of Finite State Machines*, McGraw-Hill, 1962.

Gohberg, I., Lancaster, P., and Rodman, L., *Matrices and Indefinite Scalar Products*, Birkhäuser, 1983.

Gudder, S. P., *Quantum Probability*, Academic Press, 1988.

Jacobson, N., *Basic Algebra, Vol. I*, Freeman, 1985.

O'Meara, O. T., *Introduction to Quadratic Forms*, Academic Press, 1963.

Parthasarathy, K. R., *An Introduction to Quantum Stochastic Calculus*, Birkhäuser, 1992.

Pretzel, O., *Error Correcting Codes and Finite Fields*, Oxford University Press, 1996.