

Uniform distribution of initial states: The physical basis of probability

Zhang Kechen

Department of Physics, Henan Medical University, Zhengzhou, Henan 450052, People's Republic of China

(Received 14 September 1988; revised manuscript received 14 August 1989)

For repetitive experiments performed on a deterministic system with initial states restricted to a certain region in phase space, the relative frequency of an event has a definite value insensitive to the preparation of the experiments only if the initial states leading to that event are distributed uniformly in the prescribed region. Mechanical models of coin tossing and roulette spinning and equal *a priori* probability hypothesis in statistical mechanics are considered in the light of this principle. Probabilities that have arisen from uniform distributions of initial states do not necessarily submit to Kolmogorov's axioms of probability. In the finite-dimensional case, a uniform distribution in phase space either in the coarse-grained sense or in the limit sense can be formulated in a unified way.

I. INTRODUCTION

The problem with which this paper is concerned may be illustrated by an example. When a coin is repeatedly thrown on a table, heads and tails are expected to come up with equal relative frequency regardless of the exact manner of throwing. For example, it has been verified that the probability of heads remains $\frac{1}{2}$, even if the coin is released always with its head initially facing upward. Because the initial conditions may be asymmetric, the stability of the relative frequency cannot be simply attributed to the symmetry of the coin without considering the underlying dynamics of coin tossing. Then, from a physical point of view, how is this stability possible?

Suppose the effect of external disturbances is negligible, then according to the laws of classical mechanics, the outcome of each trial is determined by the initial state of the coin (i.e., the state at the instant immediately after having been released). To throw the coin is simply to choose or set up an initial state. The phase space of the coin can be decomposed into two sets, H and T , where H consists of all phase points which, serving as initial states, will definitely lead to the occurrence of heads. Now the problem is to explain why the initial states for repetitive experiments are chosen invariably as often from set H as from set T no matter how the coin is thrown.

For the throwing of an asymmetric die, there is an additional problem: Given the mechanical properties of the die, how can one predict the probability of each face accordingly?

Such problems are readily solved by rather simple arguments. Without loss of generality, take also coin tossing as an example. Let I be the collection of all reasonable initial states for the repetitive experiments, excluding those initial states whose energy is either too small or too large, etc. It is expected that in phase space set H occupies half of the volume of set I , i.e.,

$$m(H \cap I)/m(I) = \frac{1}{2}, \tag{1.1}$$

m denoting phase volume. In repetitive experiments, if all points in I are chosen as initial states according to a probability density f , then the probability of heads is

$$p = \int_{H \cap I} f \, dm, \tag{1.2}$$

with

$$\int_I f \, dm = 1. \tag{1.3}$$

The explicit form of f is arbitrary, depending on the manner in which the coin is thrown. Strictly speaking, probability p defined by Eq. (1.2) cannot be an invariant independent of f . Nevertheless, there exists an important special case: If in the phase space, set H is distributed uniformly in fine grains within set I (cf. the distribution in Fig. 1), then

$$p \simeq \frac{1}{2} \tag{1.4}$$

holds for any function f that is not sharply peaked. In fact, the uniformness of the distribution means that when set I is decomposed into many regular shaped small cells I_1, I_2, \dots, I_n , the density of H within each cell is close to $\frac{1}{2}$, or

$$m(H \cap I_i)/m(I_i) \simeq \frac{1}{2}, \quad i = 1, 2, \dots, n \tag{1.5}$$

and the smoothness of f means that it can be approximated in each cell I_i by a certain constant f_i . Hence

$$\begin{aligned} p = \int_{H \cap I} f \, dm &\simeq \sum_i f_i m(H \cap I_i) \simeq \frac{1}{2} \sum_i f_i m(I_i) \\ &\simeq \frac{1}{2} \int_I f \, dm = \frac{1}{2}. \end{aligned}$$

Now it becomes clear that the probability of heads is insensitive to the exact manner of throwing only if set H is distributed uniformly within set I .

This argument is readily applied to the throwing of an asymmetric die. The only difference is that the density of the uniform distribution need not be $\frac{1}{2}$. As a matter of fact, the probability of each face is predicted to be equal to the corresponding density.

For later consideration, it is helpful to introduce some special terms. When repetitive experiments are performed on a deterministic system, the collection of all allowable initial states will be called the *set of initial condi-*

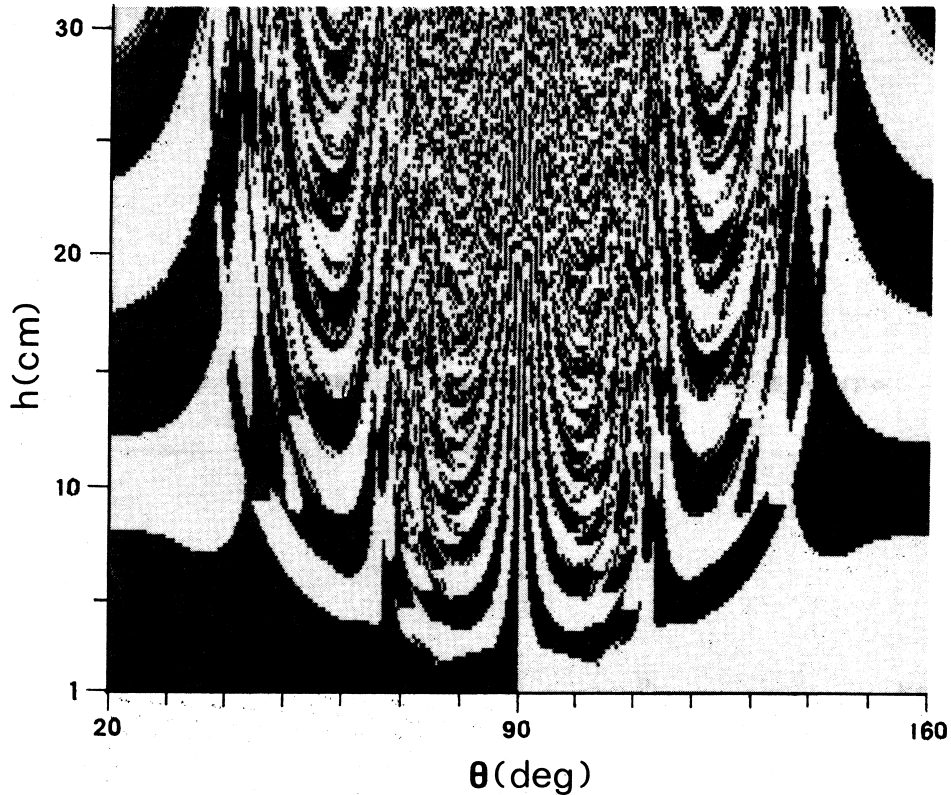


FIG. 1. Dependence of the final outcome on the initial state in the mechanical model of coin tossing. The coin is initially at rest when released above the table. Blackened regions represent the initial states leading to heads. Parameters: $r=0.6$ cm and $e=0.5$. Computation step: 0.1875 cm \times 0.4375 deg.

tions for the experiments (e.g., set I in the previous example). An event is said to have *intrinsic probability* of value D if in the phase space the initial states leading to that event are distributed uniformly within the set of initial conditions, and the density is D . The collection of those initial states (e.g., set H) can also be called an *event*.

The main idea in the preceding argument is not new. Uniform distribution was used by Poincaré to explain the equidistribution in roulette wheels and minor planets.¹ Poincaré's method of arbitrary functions was subsequently developed by Hopf, who especially realized its general importance to the foundations of probability theory.²⁻⁴ It is widely known that a deterministic system may exhibit random behavior,⁵⁻⁸ but the present topic is rarely considered.

In the present paper, this method will be revitalized. Both the physical implications and the mathematical formulation of intrinsic probability will be elaborated, and the principle that uniform distribution of initial states leads to definite probability will be elevated to an independent physical hypothesis.

II. EXAMPLES IN CLASSICAL MECHANICS

A. Coin tossing

A realistic mechanical model of coin tossing will be constructed to examine whether the initial states leading to heads (or tails) are distributed uniformly in phase space, that is, whether there exists intrinsic probability.

This model can be generalized to die throwing.

The coin, a round disk with uniform mass distribution is to be released above a plain, smooth table. The presence of air is ignored. To simplify the problem, the initial state of the coin is so chosen that the velocity of the center of mass is vertical and the angular velocity is parallel to the tabletop.⁹ This mode of motion will be automatically preserved, because all external forces experienced by the coin, namely, the force F exerted by the table and the gravity, are vertical.

In the freely falling period, the motion is simply a free fall plus a rotation of constant angular velocity. During the course of collision, by neglecting gravity and retaining force F , the dynamical equations can be written as

$$m\dot{v} = F, \quad (2.1a)$$

$$I\dot{\omega} = FrS(\theta)\cos\theta, \quad (2.1b)$$

where m , r , and v are the mass, radius, and velocity at the center of mass of the coin, respectively; $I = \frac{1}{4}mr^2$ is the moment of inertia; θ is the angle between the surface of the table and that of the coin ($\theta=0$ when the coin is parallel to the tabletop with heads up); $\omega = \dot{\theta}$; and

$$S(\theta) = -\text{sgn}(\sin\theta) = \begin{cases} 1 & \text{if } \sin\theta < 0, \\ -1 & \text{if } \sin\theta > 0, \end{cases}$$

which is necessary for Eqs. (2.1) since the coin may contact the table at either of the two opposite points of the

edge. Since a collision takes a very short time, it is assumed that the position of the coin undergoes no change during that course. Canceling F from Eqs. (2.1), a subsequent integration yields

$$S(\theta)(v' - v)\cos\theta = \frac{1}{4}r(\omega' - \omega), \quad (2.2)$$

where primed quantities refer to the instant immediately after the collision and unprimed ones to that immediately before.

The essential assumption on the collision is that the vertical component of velocity at the edge point colliding with the table attenuates a constant proportion for each collision, or

$$\frac{v' + \omega' r S(\theta) \cos\theta}{v + \omega r S(\theta) \cos\theta} = -e, \quad (2.3)$$

where e is a constant between 0 and 1 known as the coefficient of restitution. It is mentioned that this assumption is equivalent to

$$(E' - E^*) / (E - E^*) = e^2, \quad (2.4)$$

where E , E' , and E^* are the kinetic energy of the coin immediately before and after the collision and at the intermediate instant when the vertical velocity at the edge point is zero, respectively.

Combining Eqs. (2.2) and (2.3), one has

$$v' = \frac{(4 \cos^2\theta - e)v - [(1+e)rS(\theta)\cos\theta]\omega}{1 + 4 \cos^2\theta}, \quad (2.5a)$$

$$\omega' = \frac{-[4(1+e)S(\theta)\cos\theta]v + r(1-4e \cos^2\theta)\omega}{r(1 + 4 \cos^2\theta)}. \quad (2.5b)$$

This completes the model.

The state of the coin is specified by (h, θ, v, ω) , where h is the height of the center of mass. Given the initial state, the subsequent motion can be calculated piecewise from one collision to the next. The final outcome (heads or tails) can be determined whenever the kinetic energy becomes less than mgr , for the coin can no longer turn over thereafter. It should be noted that this model is unable to treat degenerate collisions. One is the "flat" collision with $\sin\theta=0$, the other is the "touch" collision with the denominator in Eq. (2.3) being zero. Although in the latter case Eqs. (2.5) predict that $v'=v$ and $\omega'=\omega$, this is a reasonable solution only if the edge point colliding with the table is accelerating upward, or

$$-g - \omega^2 r S(\theta) \sin\theta > 0.$$

Let H be the collection of all initial states leading to heads. The distribution of H in phase space is readily revealed with the aid of a computer. Figure 1 shows the distribution restricted to the plane $v = \omega = 0$.

Imagine that a machine is designed to release the coin with precisely controlled initial values of θ and h (within errors $\Delta\theta$ and Δh , respectively). The set of initial conditions $I_{\theta h}$ for this experiment, when represented in Fig. 1, is a rectangle centered at (θ, h) with sides of length $2\Delta\theta$ and $2\Delta h$. When this experiment is repeated, the long-term relative frequency of heads is expected to be around $\frac{1}{2}$ only if set H is distributed uniformly in fine grains within set $I_{\theta h}$. This requirement is likely to be satisfied if

h is sufficiently large. In fact, it may be conjectured that for any fixed $\Delta\theta$, Δh , and θ ,

$$\lim_{h \rightarrow \infty} m(H \cap I_{\theta h}) / (I_{\theta h}) = \frac{1}{2}, \quad (2.6)$$

where m denotes area on the real plane.¹⁰ Moreover, it may also be conjectured that H is distributed uniformly within any region in phase space where the energy is large enough, and within any fixed region when coefficient e is sufficiently close to 1.¹¹

This model can be generalized. Take the throwing of an asymmetric die as an example. The motion in the freely falling period can be solved by Poincot's construction.¹² During the course of collision, the dynamical equations are

$$I_{xx}\dot{\omega}_x + I_{xy}\dot{\omega}_y + I_{xz}\dot{\omega}_z = ym\dot{v}, \quad (2.7a)$$

$$I_{yx}\dot{\omega}_x + I_{yy}\dot{\omega}_y + I_{yz}\dot{\omega}_z = -xm\dot{v}, \quad (2.7b)$$

$$I_{zx}\dot{\omega}_x + I_{zy}\dot{\omega}_y + I_{zz}\dot{\omega}_z = 0, \quad (2.7c)$$

where coordinates (x, y, z) refer to the space axes with the origin at the center of mass and the z axis being vertical, and they also represent the position of the vertex that contacts the table; m is the mass; I_{xx} , I_{xy} , etc. are the moment of inertia coefficients; $(\omega_x, \omega_y, \omega_z)$ is the angular velocity; and v is the vertical component of velocity at the center of mass. The counterpart of assumption (2.3) is

$$y\omega'_x - x\omega'_y + v' = -e(y\omega_x - x\omega_y + v). \quad (2.8)$$

Integrating Eqs. (2.7) yields linear equations similar to Eq. (2.2). Putting them and Eq. (2.8) together, one can write

$$A\eta' = B\eta, \quad (2.9)$$

where $\eta = (\omega_x, \omega_y, \omega_z, v)$, η' is the vector with corresponding primed components, and A and B are 4×4 matrices. It is not difficult to verify that

$$\det A = -e \det B < 0.$$

Therefore, η' can be solved uniquely from (2.9), and the model is thus completed.

Compared with real coin tossing and die throwing, an apparent shortcoming of the present model is that the force exerted by the table has no horizontal components. Whether there exists intrinsic probability in real systems may be decided by experiments with precisely controlled initial states.

B. Roulette spinning

Here intrinsic probability associated with the final resting position of a freely spinning model roulette wheel is considered. The total torque of damping forces is assumed to depend only on the angle θ that specifies the position of the wheel and on the angular velocity $\omega = \dot{\theta}$; that is, the wheel obeys the dynamical equation of the form

$$\dot{\omega} = -F(\theta, \omega). \quad (2.10)$$

The state of this deterministic system is specified by θ and ω . For simplicity, the phase space will be represent-

ed as a real plane, although it is more appropriately represented as a cylinder, for all angles of the form $2k\pi + \theta$ correspond to the same position of the wheel.

First set

$$F(\theta, \omega) = \epsilon X(\theta)\omega, \tag{2.11}$$

where $X(\theta) > 0$ is any function of the period 2π , and $\epsilon > 0$ is a parameter. Then

$$\frac{d\omega}{dt} = \frac{d\omega}{d\theta} \omega = -\epsilon X(\theta)\omega,$$

or

$$\frac{d\omega}{d\theta} = -\epsilon X(\theta). \tag{2.12}$$

It is evident that the trajectories in phase space form a family of parallel curves, such as illustrated by Fig. 2.

Consider the event where angle θ finally lies within the sector $(2k\pi, 2k\pi + \Delta\theta)$. The collection $A_\epsilon(\Delta\theta)$ of the initial states leading to this event is a family of parallel strips in the phase space (Fig. 2). Vertically, the width of the strips is

$$\Delta\omega = \epsilon \int_0^{\Delta\theta} X(\theta) d\theta, \tag{2.13}$$

and the vertical period of them is

$$u = \epsilon \int_0^{2\pi} X(\theta) d\theta \equiv \epsilon/c. \tag{2.14}$$

The distribution of the event $A_\epsilon(\Delta\theta)$ depends on parameter ϵ . As $\epsilon \rightarrow 0$, both $\Delta\omega$ and u approach 0, while $\Delta\omega/u$ remains a constant. It is easily seen that for any area I on the θ - ω plane,

$$\lim_{\epsilon \rightarrow 0} m(A_\epsilon(\Delta\theta) \cap I) / m(I) = \Delta\omega / u \equiv D(\Delta\theta). \tag{2.15}$$

Let the set of initial conditions be any area on the θ - ω

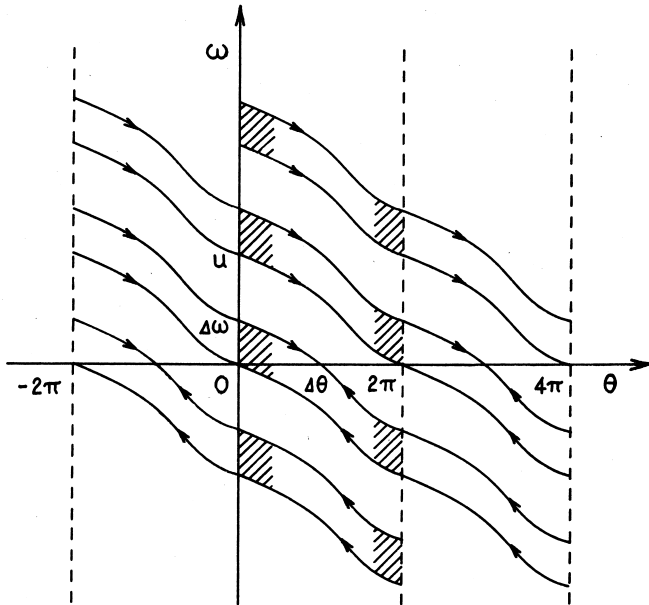


FIG. 2. Phase space of a model roulette wheel showing parallel trajectories running towards the line $\omega=0$. Partially shaded strips represent the set $A_\epsilon(\Delta\theta)$, the collection of all initial states starting from which angle θ finally lies within the sector $(2k\pi, 2k\pi + \Delta\theta)$.

plane, then event $A_\epsilon(\Delta\theta)$ has an intrinsic probability $D(\Delta\theta)$, provided that ϵ is small enough. In other words, when the wheel comes to rest angle θ has an intrinsic probability density

$$\frac{dD(\Delta\theta)}{d\Delta\theta} = cX(\theta) \quad (0 < \theta < 2\pi), \tag{2.16}$$

provided that ϵ is small enough.

Now consider a more general case

$$F(\theta, \omega) = -\epsilon X(\theta)Y(\omega). \tag{2.17}$$

Introducing a new variable Ω defined by

$$d\Omega = [Y(\omega)]^{-1} \omega d\omega, \tag{2.18}$$

one gets

$$\frac{d\Omega}{d\theta} = -\epsilon X(\theta). \tag{2.19}$$

The state of the wheel can be specified either by (θ, ω) or by (θ, Ω) , which may be regarded as two different coordinate systems of the phase space. By comparing Eq. (2.19) with Eq. (2.12), it is obvious that in the θ - Ω plane everything is identical to that in the previous case. Because intrinsic probability does not rely on the choice of coordinate systems (see Sec. III), it can be concluded that the final value of θ also has an intrinsic probability density $cX(\theta)$ as $\epsilon \rightarrow 0$. Intrinsic probability may also arise when the initial value of ω is large enough. A necessary condition is that when the parallel strips for the event $A_\epsilon(\Delta\theta)$ in the θ - Ω plane are transformed to the θ - ω plane, the image should have a vanishing vertical period as $\omega \rightarrow \infty$. By (2.18), this implies that

$$\lim_{\omega \rightarrow \infty} Y(\omega)/\omega = 0 \tag{2.20}$$

(cf. Ref. 2). In some cases, e.g., $Y(\omega) = \text{const}$, it is easily verified that as $\omega \rightarrow \infty$ the final value of θ also has an intrinsic probability density $cX(\theta)$.²

C. Mixing system and equal a priori probability hypothesis

The state of an isolated mechanical system evolves on the energy surface X in phase space. The time evolution is prescribed by a group of transformation $\{T_t\}$ on X with $T_s T_t = T_{s+t}$ for all real numbers s and t . The system is called (strong) mixing if for all measurable sets A and B on X ,

$$\lim_{t \rightarrow \infty} m(B \cap T_{-t} A) = m(B)m(A), \tag{2.21}$$

where m is an invariant measure on X with $m(X) = 1$.^{2,13}

The equal a priori probability hypothesis is generally considered a basic postulate of equilibrium statistical mechanics.^{14,15} It asserts that in thermodynamic equilibrium the state of the mechanical system can be in any measurable set A on X with the probability

$$p = m(A). \tag{2.22}$$

It is easily shown that this hypothesis is valid if and only if the system is mixing. Suppose an initial state x_0 is arbitrarily chosen from the energy surface X to see whether the state at time t is in set A . Obviously $T_{-t}A$ is the collection of all initial states leading to this outcome. Here t is sufficiently large so that the system is expected to be in thermodynamic equilibrium at time t . Hypothesis (2.22) is practically meaningless unless probability p does not depend on the manner in which x_0 is chosen. Therefore, the event $T_{-t}A$ must have an intrinsic probability of value $m(A)$, that is, it must be distributed uniformly on the energy surface X such that the density within any measurable set B on X , namely,

$$m(B \cap T_{-t}A)/m(B),$$

should be close to $m(A)$, provided that t is large enough. As $t \rightarrow \infty$, this approximate relation changes to Eq. (2.21). The arbitrariness of A and B implies that the system is mixing. Conversely, if the system is mixing, the probability assigned by (2.22) is justified since it is obviously intrinsic.

III. GENERAL PROPERTIES OF INTRINSIC PROBABILITY

The formulation of intrinsic probability is particularly simple when a limit parameter is associated either with the event or with the set of initial conditions. Like Eqs. (2.15) and (2.21), in the former case, event A_a is said to have intrinsic probability D as parameter a is sufficiently large if

$$\lim_{a \rightarrow \infty} m(A_a \cap J)/m(J) = D \quad (3.1)$$

holds for any volume J within the set of initial conditions I . In the latter case the formulation is essentially the same, such as Eq. (2.6). It follows from Eq. (3.1) that for any integrable function φ ,

$$\lim_{a \rightarrow \infty} \int_{A_a \cap I} \varphi dm = D \int_I \varphi dm. \quad (3.2)$$

Particularly, let φ be the probability density for the choice of initial states, then

$$p = D, \quad (3.3)$$

where p is the probability of event A_a for sufficiently large a .

In the limit notation it is rigorously true that intrinsic probability does not rely on the choice of a coordinate system for the phase space. (This fact has been used in Sec. II B.) In Eq. (3.1) m represents phase volume with respect to a certain coordinate system. Let m' represent phase volume in a new coordinate system, then there exists a function φ such that

$$dm' = \varphi dm. \quad (3.4)$$

Thus

$$m'(A_a \cap J)/m'(J) = \left[\int_{A_a \cap J} \varphi dm \right] / \left[\int_J \varphi dm \right] \rightarrow D \quad (a \rightarrow \infty).$$

This means that in the new coordinate system A_a also has an intrinsic probability D .

The notion of "uniform distribution in fine grains" is not needed for the limit formulation of intrinsic probability, but it is hardly avoidable because the limit parameters are actually fixed for real repetitive experiments. As is shown in Appendix A, the term mentioned above may be quantified by the *uniformness scale*, which is intended to describe the characteristic length of the "fine grains" in phase space, and is at the same time compatible with the limit formulation. If the set of initial conditions I is an interval of length L and event A consists of n smaller intervals within I , the uniformness scale can be roughly defined as $u = L/n$. Arbitrarily choose a point (initial state) x from I and then choose the point $x' = x + \epsilon$, where ϵ is a small constant. u can be obtained from

$$P_{+-} = P_{-+} = \epsilon/u, \quad (3.5)$$

where P_{+-} is the probability for $x \in A$ and $x' \in I \setminus A$, etc., and ϵ is small enough. It is reasonable to demand that

$$\epsilon/u < D(1-D), \quad (3.6)$$

where $D = P_{++} + P_{+-}$ is the density of A within I . An equivalent form of (3.6) is

$$P_{+-}P_{-+} < P_{++}P_{--}. \quad (3.7)$$

If set I is of higher dimension, this method can also be used to define the uniformness scale with respect to one variable.

If an event has intrinsic probability D , one can predict that

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n a_i/n = D, \quad (3.8)$$

where $\{a_i\}$ is the result sequence of repetitive experiments, that is, $a_i = 1$ if the event has occurred in the i th trial, and $a_i = 0$ if otherwise. Furthermore, $\{a_i\}$ must be completely equidistributed,^{16,17} namely, given any number k ($k = 2, 3, \dots$) and any binary sequence of length k containing s ones and $k-s$ zeros, let N_n be the number of times this sequence appeared in the first n terms of the k -dimensional sequence

$$(a_1, a_2, \dots, a_k), (a_2, a_3, \dots, a_{k+1}), \dots, \quad (3.9)$$

then

$$\lim_{n \rightarrow \infty} N_n/n = D^s(1-D)^{k-s}. \quad (3.10)$$

Equation (3.10) follows from the fact that intrinsic probabilities of noninteracting systems are independent; that is, if in system i ($i = 1, 2, \dots, n$) event $A_a^{(i)}$ has intrinsic probability D_i with respect to the set of initial conditions I_i as $a \rightarrow \infty$, then in the system composed of systems 1 to n without interactions event $A_a^{(1)} \times A_a^{(2)} \times \dots \times A_a^{(n)}$ has intrinsic probability

$$D = D_1 D_2 \dots D_n \quad (3.11)$$

with respect to $I_1 \times I_2 \times \dots \times I_n$ as $a \rightarrow \infty$.² Since re-

petitive experiments performed either on a single system or on several identical systems by turns should yield result sequences of the same statistical properties, each term in sequence (3.9) can be considered as being generated by a single trial of an imaginary system composed of k identical systems. Applying Eq. (3.11) to the composite system, one gets Eq. (3.10).

The case without intrinsic probability needs some remarks. A simple example is the standing of a razor blade on a table. When this experiment is repeated, the relative frequency of each sided is determined by the method of the experiment and even the psychology of the experimenter, rather than by the physics of the blade-take system itself. Moreover, there may exist correlations in the result sequence. In general, whenever Eq. (3.10) is violated, for instance, the average length of consecutive 1's in $\{a_i\}$ is significantly different from $(1-D)^{-1}$, one can conclude that the event has no intrinsic probability.

Sometimes the probability of the event that has no intrinsic probability can vary only within a certain range. Consider the example in Fig. 3, where

$$A_n = T^{-n}L, \tag{3.12a}$$

$$B_n = (L \cap T^{-n}L) \cup (R \cap T^{-n}R). \tag{3.12b}$$

Here, $L = [0, \frac{1}{2}) \times [0, 1)$, $R = [\frac{1}{2}, 1) \times [0, 1)$, and T is the baker's transformation on the phase space $X = [0, 1) \times [0, 1)$, which transforms (x, y) to $(2x - [2x], \frac{1}{2}y + \frac{1}{2}[2x])$ with $[]$ denoting the integer part of a number.^{2,13} Suppose one is to mark a point chosen arbitrarily from X , then A_n is the event where the marked point finally appears in L under transformation T^n , and B_n is the event where the point is either initially chosen from L and finally appears also in L , or initially and finally both in R . If n is large enough, both A_n and B_n have intrinsic probability while $A_n \cup B_n$ has not. However, given any probability density for the choice of the marked point, the probability of $A_n \cup B_n$ is necessarily between $\frac{1}{2}$ and 1.¹⁸

IV. PHYSICAL ORIGIN OF PROBABILITY

Recall the naive meaning of probability. For repetitive experiments performed on a physical system, an event is said to have probability p if its occurrence in each individual trial cannot be predicted with certainty while its

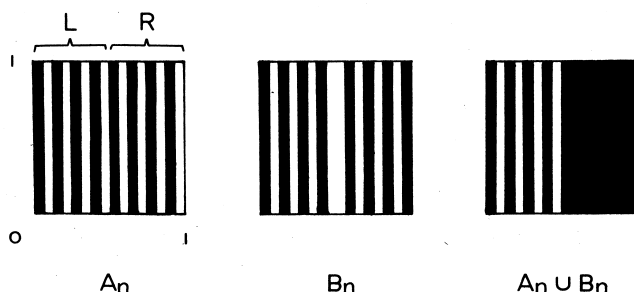


FIG. 3. Both A_n and B_n have intrinsic probability $\frac{1}{2}$, while the probability of $A_n \cup B_n$ is between $\frac{1}{2}$ and 1 without definite value. Blackened regions represent the events indicated below with $n = 3$ [see Eqs. (3.12)].

long-term relative frequency approaches p . Since this probabilistic behavior is first of all a physical phenomenon, there should be no surprise if its existence relies on some physical condition.

In the case where the physical system is deterministic, intrinsic probability just furnishes the condition. On the one hand, intrinsic probability, or the uniform distribution of initial states, implies sensitive dependence of the final outcome on the initial state, and hence the practical unpredictability of the outcome. On the other hand, intrinsic probability ensures the stability of long-term relative frequency. In other words, an event has probability p in the above-mentioned naive sense if and only if it has intrinsic probability of the value p .

It should be noted that this principle does not rely on probability theory, although it has been arrived at with the aid of a normalized function, which is referred to as probability density for the choice of initial states in the previous sections. This interpretation is actually unnecessary. As shown in Appendix B, the use of this function may be justified merely by the ergodic property of the whole process of repetitive experiments. In fact, even if the choice of initial states cannot be described by a single density function, this principle may still hold.

This principle may be simply regarded as a physical hypothesis, which relates the very existence as well as the value of probability with the deterministic dynamics, and whose validity is subject to experimental verification. This hypothesis is rather general. For example, once it is adopted, the equal *a priori* probability hypothesis should no longer be regarded as an independent postulate (cf. Sec. II C). This hypothesis also provides a starting point to justify the laws of probability. For example, the formulation (3.11) of independent events follows naturally from the definition of intrinsic probability.

It is interesting to note that intrinsic probability in general does not submit to Kolmogorov's mathematical axioms of probability.¹⁹ This is because the fact that two events both have intrinsic probability does not imply that their union also has intrinsic probability (see the example in Fig. 3). Let \mathcal{J} be the collection of all events that have intrinsic probability in the sense of Eq. (3.1). It is straightforward to verify that (i) if $A \in \mathcal{J}$, then $A^c \in \mathcal{J}$ and $D(A) + D(A^c) = 1$; and (ii) if $A \in \mathcal{J}$, $B \in \mathcal{J}$, and $A \cap B = \emptyset$, then $A \cup B \in \mathcal{J}$ and $D(A \cup B) = D(A) + D(B)$. Here D represents the value of intrinsic probability, and the subscript of limit parameter is omitted. It follows from (i) and (ii) that for any $A \in \mathcal{J}$ and $B \in \mathcal{J}$, $A \cup B \in \mathcal{J}$ if and only if $A \cap B \in \mathcal{J}$. It is noteworthy that \mathcal{J} together with the set function $D: \mathcal{J} \rightarrow [0, 1]$ happens to be a finitely additive quantum-mechanical probability space.²⁰

ACKNOWLEDGMENTS

I would like to thank Professor Glenn Shafer for recommending to me the literature of Poincaré's method of arbitrary functions.

APPENDIX A: UNIFORMNESS SCALE

Consider the distribution of event A within the set of initial conditions I , where I is a bounded and connected

open set in finite-dimensional Euclidean space, and I and $A \cap I$ are measurable sets of finite measure with respect to the Lebesgue measure m . $\xi = \{I_1, I_2, \dots, I_n\}$ is called a partition of I if the I_i 's are disjoint measurable subsets of I with nonzero measures and their union is I . ξ is called a uniform partition if it satisfies the additional condition

$$m(A \cap I_i)/m(I_i) = D \equiv m(A \cap I)/m(I) \quad (\text{A1})$$

for all I_i . The uniformness scale of event A is defined by

$$u = \inf\{\text{diam}\xi: \xi \text{ is a uniform partition}\}, \quad (\text{A2})$$

where $\text{diam}\xi = \max\{\text{diam}I_1, \dots, \text{diam}I_n\}$ and $\text{diam}I_i$ is the supremum of the distance between any two points in I_i . Except for the trivial cases $D=0$ or 1 , it is always true that $u > 0$.²¹ In the case where I is an interval in the real axis and A consists of smaller intervals of equal length distributed with period T within I , $u = \max\{D, 1-D\}T$. When set A is slightly modified, u remains of the same order.

Consider the probability of event A defined by

$$p = \int_{A \cap I} f \, dm. \quad (\text{A3})$$

It is assumed that the probability density f is a measurable function on I , and there exists a constant M such that

$$\sum_i \int_{I_i} |f - f_i| \, dm \leq M \text{diam}\xi \quad (\text{A4})$$

holds for any partition $\xi = \{I_1, I_2, \dots, I_n\}$, where

$$f_i = \frac{1}{m(I_i)} \int_{I_i} f \, dm. \quad (\text{A5})$$

For example, if I is an interval of finite length and f is of bounded variation, one can let M equal the total variation.²² It is not difficult to prove that

$$|p - D| \leq uM. \quad (\text{A6})$$

The error can also be estimated by other methods.^{3,23}

Suppose the distribution of event A_a on phase space depends on parameter a , and the limit

$$D = \lim_{a \rightarrow \infty} m(A_a \cap I)/m(I) \quad (\text{A7})$$

exists ($D \neq 0, 1$), then A_a has intrinsic probability D in the sense of Eq. (3.1) if and only if

$$\lim_{a \rightarrow \infty} u_a = 0, \quad (\text{A8})$$

where u_a is the uniformness scale of A_a . The proof is sketched as follows. Note that for any $\epsilon > 0$ one can find a partition $\xi = \{I_1, I_2, \dots, I_n\}$ such that

$$\text{diam}(I_i \cup I_{i+1}) < \epsilon, \quad i = 1, 2, \dots, n-1.$$

If A_a has intrinsic probability, the density of A_a within each I_i can be made arbitrarily close to the overall density D_a in I by fixing parameter a with a sufficiently large value. The uniform partition $\xi' = \{I'_1, I'_2, \dots, I'_n\}$ with respect to A_a is easily constructed from ξ such that

$I'_i \subset I_i \cup I_{i+1}$ ($i = 1, 2, \dots, n-1$). Since $\text{diam}\xi' < \epsilon$, it follows that $u_a \rightarrow 0$. On the other hand, if $u_a \rightarrow 0$, employing a uniform partition ξ with $\text{diam}\xi < 2u_a$, one can prove that

$$|m(A_a \cap B_r) - D_a m(B_r)| \leq m(B_r) - m(B_{r'}) \rightarrow 0 \quad (a \rightarrow \infty),$$

where B_r and $B_{r'}$ are concentric balls contained in I of radii r and $r' = r - 2u_a$. This implies that A_a has intrinsic probability.

Finally, it is mentioned that the difference ΔD between the density of event A within an arbitrary ball B_r and the overall density D submits to

$$|\Delta D| \leq 1 - (1 - u/r)^s \sim su/r, \quad (\text{A9})$$

where s is the dimension of set I .

APPENDIX B: DENSITY FOR CHOICE OF INITIAL STATES

For concreteness, take coin tossing as an example. Suppose a machine is designed to release the coin above a table with the set of initial conditions I . Whenever the coin settles down, the machine automatically picks it up and starts the next trial. In the ideal situation without external disturbances, the composite system of the coin and the machine can be assumed to be deterministic. Its state $x^* = (x, x')$ is specified by state x of the coin and state x' of the machine. It is convenient to write $x = \pi x^*$ and $x' = \pi' x^*$. x^* is called a ready state if it is at the instant that the coin is just to be released. Given any ready state x^* , the ready state for the next trial is determined by x^* and is denoted by Tx^* . Let x^* be the ready state for the first trial, then

$$x_k = \pi T^{k-1} x^*, \quad k = 1, 2, \dots,$$

is the initial state of the coin for the k th trial.

The use of density function f for the choice of initial states is justified if for any measurable subset J of I

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \varphi_J(x_k) = \int_J f \, dm,$$

where φ_J is the characteristic function of J and m is the measure on I . Let I^* be the collection of all ready states, then $\pi I^* = I$. Suppose m' is the measure on $I' = \pi' I^*$ and measure $m^* = m \times m'$ is invariant with respect to T with $m^*(I^*) = 1$. If the abstract dynamical system $(I^*, \mathcal{A}^*, m^*, T)$ is ergodic, that is,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n g(T^{k-1} x^*) = \int_{I^*} g \, dm^*$$

holds for any measurable function g on I^* and almost every $x^* \in I^*$,¹³ then it is straightforward to verify that

$$f(x) = \int_{I'_x} dm'$$

is the desired density function with

$$I'_x = \{\pi' x^*: x^* \in I^*, \pi x^* = x\}.$$

- ¹H. Poincaré, *Calcul des Probabilités* (Gauthier-Villars, Paris, 1912), p. 148; *The Foundations of Science* (Science Press, New York, 1912), p. 163.
- ²E. Hopf, *J. Math. and Phys.* **13**, 51 (1934); **14**, 4 (1935); *Math. Ann.* **114**, 161 (1937).
- ³P. Diaconis and E. Engel, *Stat. Sci.* **1**, 171 (1986).
- ⁴J. von Plato, *Br. J. Philos. Sci.* **34**, 37 (1983).
- ⁵J. Moser, *Stable and Random Motions in Dynamical Systems* (Princeton University Press, Princeton, 1973).
- ⁶D. Ruelle and F. Takens, *Commun. Math. Phys.* **20**, 167 (1971); J. P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985).
- ⁷J. Ford, *Phys. Today* **36**, (4), 40 (1983).
- ⁸S. Wolfram, *Phys. Rev. Lett.* **55**, 449 (1985).
- ⁹Yue Z.-y. and Zhang B., *Appl. Math. Mech.* **6**, 193 (1985).
- ¹⁰For coin tossing without collision, a similar statement in the v - ω plane is proved analytically in J. B. Keller, *Am. Math. Mon.* **93**, 191 (1986).
- ¹¹Note that an infinitesimal change of initial state may lead to a different outcome only if in one of the collisions the angle θ is in the vicinity of the critical values defined by $\sin\theta=0$, $\cos\theta=0$, and $\cos\theta=\sqrt{e}/2$ or $1/2\sqrt{e}$, at which the coefficients of v and ω in Eqs. (2.5) change signs. [The latter two correspond to $\theta=45^\circ(135^\circ)$, $69^\circ(111^\circ)$ in the example in Fig. 1.] For larger initial energy or e closer to 1, the change of initial state, which suffices to cause a different outcome, is likely to be smaller, because the coin tends to bounce more times before it is unable to turn over and collisions at critical angles are more likely to occur.
- ¹²H. Goldstein, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1980), p. 205.
- ¹³V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968); I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory* (Springer-Verlag, New York, 1981); P. Walters, *An Introduction to Ergodic Theory* (Springer-Verlag, New York, 1982).
- ¹⁴R. C. Tolman, *The Principle of Statistical Mechanics* (Oxford University Press, Oxford, England, 1938).
- ¹⁵A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949); I. E. Farquhar, *Ergodic Theory in Statistical Mechanics* (Wiley, London, 1964).
- ¹⁶J. N. Franklin, *Math. Computation* **17**, 28 (1963).
- ¹⁷D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms* (Addison-Wesley, Reading, MA, 1969), p. 127.
- ¹⁸In general, if the limit $D(x) = \lim_{r \rightarrow 0} \lim_{a \rightarrow \infty} m[A_a \cap B_r(x)]/m[B_r(x)]$ exists for all x in the set of initial conditions I , where $B_r(x)$ is the ball of radius r centered at x , then the probability of event A_a is $p = \lim_{a \rightarrow \infty} \int_{A_a \cap I} f dm = \int_I f(x) D(x) dx$, and $\inf_{x \in I} D(x) \leq p \leq \sup_{x \in I} D(x)$.
- ¹⁹A. N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea, New York, 1956); A. N. Shiriyayev, *Probability* (Springer-Verlag, New York, 1984).
- ²⁰P. Suppes, *Philos. Sci.* **33**, 14 (1966).
- ²¹This is a consequence of Lebesgue's density theorem, see, e.g., J. C. Oxtoby, *Measure and Category*, 2nd ed. (Springer-Verlag, New York, 1980), p. 16.
- ²²See, e.g., G. B. Folland, *Real Analysis: Modern Techniques and Applications* (Wiley, New York, 1984), p. 95.
- ²³Note an interesting case. Suppose that event A consists of intervals of length L distributed on $I = (-\infty, \infty)$ with period T , and that the Fourier transform of function f vanishes outside the interval $[-B/2, B/2]$, then $p = L/T$ holds exactly as long as $TB < 2$.

Errata

Erratum: Uniform distribution of initial states: The physical basis of probability [Phys. Rev. A 41, 1893 (1990)]

Zhang Kechen

Inadvertently, some corrections were not incorporated in the final production stages of this paper.

(1) Equation (2.6) should read

$$\lim_{h \rightarrow \infty} m(H \cap I_{\theta h}) / m(I_{\theta h}) = \frac{1}{2}. \quad (2.6)$$

(2) Equation (2.17) should read

$$F(\theta, \omega) = \epsilon X(\theta) Y(\omega). \quad (2.17)$$

(3) On p. 1898, left-hand column, the third sentence of the first complete paragraph should read as follows:

When this experiment is repeated, the relative frequency of each side is determined by the method of the experiment and even the psychology of the experimenter, rather than by the physics of the blade-table system itself.

(4) Figure 1 is reproduced below using a diagram with better contrast than originally presented.

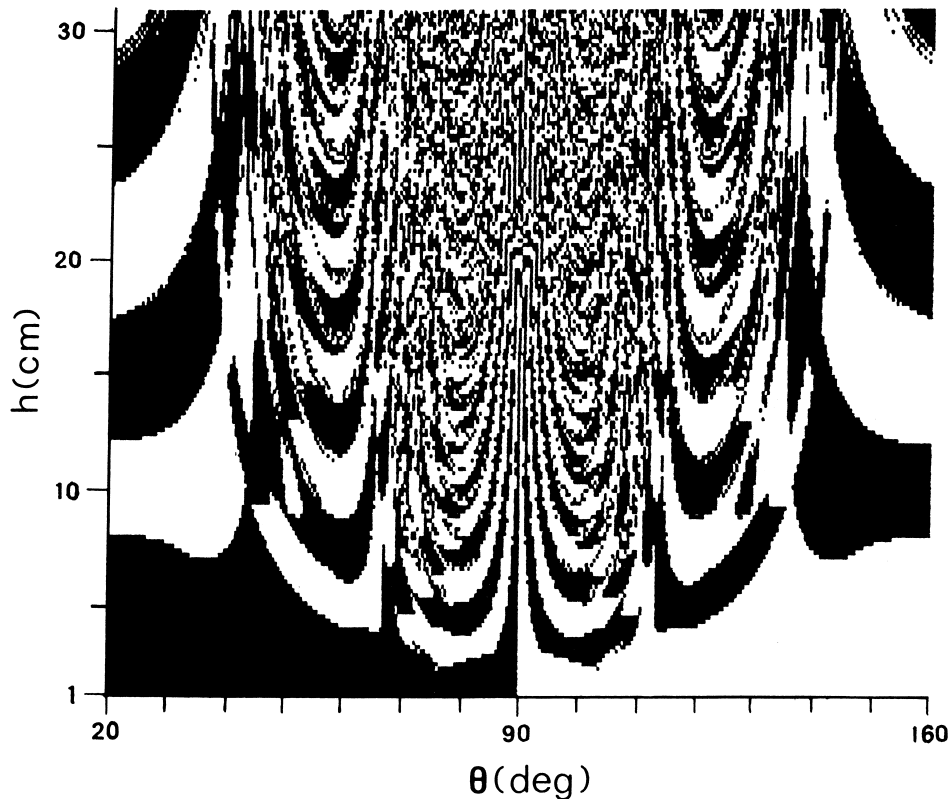


FIG. 1. Dependence of the final outcome on the initial state in the mechanical model of coin tossing. The coin is initially at rest when released above the table. Blackened regions represent the initial states leading to heads. Parameters: $r = 0.6$ cm and $e = 0.5$. Computation grid: 0.1875 cm \times 0.4375 deg.