A NEW DEDUCTION OF MAXWELL'S LAW OF VELOCITY DISTRIBUTION

A. Rényi (Budapest)

INTRODUCTION

At the time when the foundations of classical statistical mechanics were laid down, the theory of probability was not yet enough developed, to serve as a safe basis on which statistical mechanics could be built up. This possibility was given only after the theory of probability has been developed by A. N. Kolmogoroff [1], into an exact mathematical theory. A. Ja. Khintchine [2] was the first who recognized the necessity as well as the possibility of building statistical mechanics on the basis of modern probability theory. He fulfilled this task not only with respect to the classical statistics, but also for quantum statistics [3].

The theory of Khintchine is in many respect clearly superior to former theories, and solves with success also those serious formal difficulties which arise from the fact, that in statistical mechanics unbounded measures play an important role, and such measures can not be interpreted as probability distributions in the theory of Kolmogoroff. As a matter of fact, according to Kolmogoroff, a probability space is defined as a triple $(S, \mathbf{A}, P)$ where S is an abstract space, whose elements are the elementary events, $\mathbf{A}$ a σ-algebra of subsets of S, called events and $P=P(A)$ (the probability of the event A) a measure on $\mathbf{A}$, normed by the condition $P(S)=1$. Thus unbounded measures can not be interpreted as probability measures, for instance it has no sense to speak about a probability distribution, which is uniform in the whole n-dimensional Euclidean space E_n. But exactly such a distribution in phase-space is suggested by Liouville's theorem on the invariance of volume in phase-space with respect to the natural transformations of this space during the motion of the mechanical system considered.

The same difficulty presents himself also in other domains of application of probability theory, e. g. in quantum mechanics, integral geometry, in some problems of mathematical statistics, etc. In all these disciplines, there arise in a natural way such problems, which can be solved only in a round-about way within the frames of the theory of Kolmogoroff, owing to the fact that they lead to unbounded measures. To solve this difficulty, the author of the present paper developed recently [4] a new axiomatic theory of probability, in which unbounded measures are admitted and give rise to probability distributions. In this theory, which is a natural generalization of that of Kolmogoroff*, conditional probability is taken as the fundamental concept.

In § 1. we give a short sketch of the new theory and in § 2. we apply this theory to give a new and simple deduction of Maxwell's law of velocity distribution, from which the reader may judge the advantage of the new theory for statistical mechanics.

§ 1. A new axiomatic theory of probability

In what follows if A and B are sets, we denote by $A+B$ the set of those elements which belong to at least one of the sets A and B, and by AB the set of those elements which belong to both of the sets A and B. The empty set will be denoted by O, the subset of B consisting of those elements which do not belong to A will be denoted by $B-A$. If a is an element of the set A, this will be denoted by $a \in A$. If a does not belong to the set A, this will be denoted by $a \bar{\in} A$

We start from an arbitrary set S, the set of elementary events; the elements of S will be denoted by the letter a. Let $\mathbf{A}$ denote a σ-algebra of subsets of S; the subsets of S which are elements of $\mathbf{A}$ will be denoted by capital letters $A, B, C, \ldots$, and called (random) events. Let us suppose further that an arbitrary non empty subset $\mathbf{B}$ of $\mathbf{A}$ is given; and finally, that a set function $P(A \mid B)$ of two set-variables is defined for $A \in \mathbf{A}$ and $B \in \mathbf{B}$; $P(A \mid B)$ will be called the conditional probability of the event A, with respect to the event B. As the conditional probability of the event A, with respect to the event B is defined if and only if B, belongs to $\mathbf{B}$, $\mathbf{B}$ may be called the set of possible conditions. We suppose that the set function $P(A \mid B)$ satisfies the following axioms:

Axiom I. $P(A \mid B) \geq 0$ if $A \in \mathbf{A}$ and $B \in \mathbf{B}$ further $P(B \mid B) = 1$ if $B \in \mathbf{B}$.

Axiom II. For any fixed $B \in \mathbf{B}$, $P(A \mid B)$ is a measure (by other words a countably additive set function) of A, i. e. if $A_n \in \mathbf{A}$ $(n=1,2,\ldots)$ and $A_j A_k = O$ for $j \neq k$, $j,k=1,2,\ldots$ we have

$$P\left(\sum_{n=1}^{\infty} A_n \mid B\right) = \sum_{n=1}^{\infty} P(A_n \mid B).$$

Axiom III. If $A \in \mathbf{A}$, $B \in \mathbf{A}$, $C \in \mathbf{B}$ and $BC \in \mathbf{B}$, we have

$$P(A \mid BC)\, P(B \mid C) = P(AB \mid C).$$

* The author has been informed, that the idea of such a generalization of his theory has been pointed out in a lecture some years ago by Kolmogoroff himself, but he did not publish his ideas on the subject.

If the Axioms I—III. are satisfied, we shall call the set S, the σ-algebra $\boldsymbol{A}$ of subsets of S, the subset $\boldsymbol{B}$ of $\boldsymbol{A}$ and the set function $P(A \mid B)$, together a conditional probability space and denote it for the sake of brevity by $\boldsymbol{P} = [S, \boldsymbol{A}, \boldsymbol{B}, P(A \mid B)]$. It is easy to see that it follows from our axioms that $P(A \mid B) = P(AB \mid B)$ and that $P(A \mid B) \leq 1$ for all $A \epsilon \boldsymbol{B}$ and $B \epsilon \boldsymbol{B}$, further that $P(O \mid B) = 0$.

If $\boldsymbol{P} = [S, \boldsymbol{A}, \boldsymbol{B}, P]$ is a conditional probability space and C is an arbitrary fixed element of $\boldsymbol{B}$, putting $P_c(A) = P(A \mid C)$ clearly $[S, \boldsymbol{A}, P_c]$ will be a probability space in the sense of Kolmogoroff. Thus a conditional probability space is nothing else than a set of ordinary probability spaces, which are connected with each other by Axiom III. This connection is such that it is in conformity with the usual definition of conditional probability. Namely if we put $P_c(A) = P(A \mid C)$ for $A \epsilon \boldsymbol{A}$, with $C \epsilon \boldsymbol{B}$ fixed, and define the conditional probability $P^*(A \mid B)$ for a $B \epsilon \boldsymbol{B}$ for which $P_c(B) > 0$ as usual in the theory of Kolmogoroff, by $P^*(A \mid B) = \frac{P_c(AB)}{P_c(A)}$, we have by Axiom III. $P^*(A \mid B) = P(A \mid BC)$.

If $\xi = \xi(a)$ denotes a real-valued function defined for $a \epsilon S$ which is measurable with respect to $\boldsymbol{A}$, i. e. if A_x denotes the set of those $a \epsilon S$ for which $\xi(a) < x$ we have $A_x \epsilon \boldsymbol{A}$ for all x we shall call ξ a random variable on $[S, \boldsymbol{A}, \boldsymbol{B}, P]$. Vector valued random variables are defined similarly.

An important class of conditional probability spaces is obtained as follows:

Let us choose for S the n-dimensional Euclidean space E_n, for $\boldsymbol{A}$ the set of all measurable subsets of E_n, let $f(x)$ where $x = (x_1, x_2, \ldots, x_n)$, be a non-negative measurable function on E_n and for $\boldsymbol{B}$ choose the set of those subsets $B \epsilon \boldsymbol{A}$ for which

$$\Phi(B) = \int_B f(x) dx$$

(where dx denotes $dx_1 dx_2 \ldots dx_n$) is finite and positive. Putting

$$P(A \mid B) = \frac{\Phi(AB)}{\Phi(B)}$$

for $A \epsilon \boldsymbol{A}$ and $B \epsilon \boldsymbol{B}$, we obtain a conditional probability space $[S, \boldsymbol{A}, \boldsymbol{B}, P]$.

Especially when $f(x) \equiv 1$, we obtain a conditional probability space, for which $P(A \mid B) = \frac{m_n(AB)}{m_n(B)}$, where m_n denotes the n-dimensional Lebesgue-measure; in this case $\boldsymbol{B}$ consists of all measurable sets which have a positive and finite Lebesgue-measure. The conditional probability space thus obtained will be called the uniform conditional probability space in E_n.

As $[S, \boldsymbol{A}, P(A \mid C)]$ is for any fixed $C \epsilon \boldsymbol{B}$ a probability field in the sense of the theory of Kolmogoroff, any theorem of the ordinary probability theory remains valid in the new theory when ordinary probabilities, distributions, mean values, independence, etc. are replaced by conditional probabilities, conditional distributions, conditional mean values, conditional independence, etc. with respect to the same $C \in \boldsymbol{B}$.

If ξ is a random variable and A_u^v denotes the set consisting of those elements $a \in S$ for which $u \leqq \xi(a) < v$ and if $A_u^v \in \boldsymbol{B}$ $(u < v)$, then the conditional probabilities $P(x \leqq \xi < y \mid u \leqq \xi < v)$ can be considered for $u \leqq x < y \leqq v$ and thus ξ generates a conditional probability space on the real axis R, as the space of elementary events, the σ-algebra $\boldsymbol{A}_\xi$ being the set of Borel subsets of R, and $\boldsymbol{B}_\xi$ consisting of the intervals (u, v) for which $A_u^v \in \boldsymbol{B}$.

This conditional probability space will be called the conditional probability space generated by ξ on the real axis.

Let $F(x)$ denote a non-decreasing function of x which is continuous to the right for $-\infty < x < +\infty$. If the set $u \leqq \xi(a) < v$ belongs to $\boldsymbol{B}$, if $F(v) - F(u) > 0$ and we have for any subinterval (x, y) of such an interval (u, v)

$$(1) \qquad P(A_x^y \mid A_u^v) = \frac{F(y) - F(x)}{F(v) - F(u)}, \quad (u \leqq x < y \leqq v),$$

we shall call $F(x)$ the distribution function of ξ; the function $F(x)$ is not uniquely determined, as together with $F(x)$ the function $aF(x) + b$ where $a > 0$ and b are arbitrary constants, is also a distribution function of ξ; but as $F(x)$ will be used only to calculate the conditional probabilities (1), this will never lead to a misunderstanding. If the distribution function $F(x)$ of ξ is absolutely continuous, and $F'(x) = f(x)$, we shall call $f(x)$ the density function of ξ; clearly $f(x)$ is determined only up to a positive constant factor.

If $f(x)$ is the density function of ξ, we have

$$(2) \qquad P(A_x^y \mid A_u^v) = \frac{\int\limits_x^y f(t)\,dt}{\int\limits_u^v f(t)\,dt}$$

for $u \leqq x < y \leqq v$ provided that the denominator at the right of (2) is positive. If $f(x) = 1$ $(-\infty < x < +\infty)$, we shall say that the distribution of ξ is uniform in $(-\infty, +\infty)$.

Now we introduce the notion of the regularity of the distribution of random variables. If $\xi_1, \xi_2, \ldots, \xi_n$ are random variables on a conditional probability space $P = [S, \boldsymbol{A}, \boldsymbol{B}, P]$ and for any point $x = (x_1, \ldots, x_n)$ of E_n the set B_x, defined by the conditions $\xi_k = x_k$ $(k = 1, 2, \ldots, n)$ belongs to $\boldsymbol{B}$, further for any interval I defined by $a_k \leqq x_k < b_k$, $k = 1, 2, \ldots, n$, B_I denotes the set determined by the inequalities $a_k \leqq \xi_k < b_k$, and there exists a non-negative measurable function $f(x)$ of the variable $x = (x_1, x_2, \ldots, x_n)$ such that for any $A \in \boldsymbol{A}$ and any interval I for which $\int\limits_I f(x)\,dx > 0$ we have $B_I \in \boldsymbol{B}$ and

(3) $$P(A \mid B_t) = \frac{\int_\gamma P(A \mid B_x) f(x) dx}{\int_\gamma f(x) dx}$$

we say that the distribution of the random variables $\xi_1, \ldots, \xi_n$ is regular.

Especially if $f(x) = f_1(x_1) \ldots f_n(x_n)$, the random variables $\xi_1, \ldots, \xi_n$ are independent with respect to any $B_t \in \boldsymbol{B}$ and $f_k(x)$ is the density function of ξ_k.

If further $f_k(x) = 1$ for $-\infty < x < +\infty$ we say that the joint distribution of the variables ξ_k is uniform and regular in E_n.

Let us suppose that $\xi_1, \xi_2, \ldots, \xi_n$ are independent positive random variables on the conditional probability space $\boldsymbol{P}$ and that their distribution is regular. Let us denote by $f_k(x)\,(x>0)$ the density function of $\xi_k\,(k=1, 2, \ldots, n)$.

Let us calculate the density function of the random variable $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$. If J_T is the interval $0 \leq x_k < T\,(k = 1, 2, \ldots, n)$, B_T the set for which $0 \leq \xi_k < T$, and A_z the set defined by $\zeta_n < z$, we have by the definition of regularity

(4) $$P(A_z \mid B_T) = \frac{\displaystyle\int\limits_{\sum_{k=1}^{n} x_k < z,\; 0 \leq x_k < T} \prod_{k=1}^{n} f_k(x_k) dx_k}{\displaystyle\prod_{k=1}^{n} \left(\int_0^T f_k(x) dx \right)}$$

Now let us suppose that $0 < z_1 < z_2 < T$; in this case we have

$$P(A_{z_1} \mid A_{z_2}) = \frac{P(A_{z_1} \mid B_T)}{P(A_{z_2} \mid B_T)}$$

and thus

(5) $$P(A_{z_1} \mid A_{z_2}) = \frac{\displaystyle\int\limits_{\sum_{k=1}^{n} x_k < z_1} \prod_{k=1}^{n} f_k(x_k) dx_k}{\displaystyle\int\limits_{\sum_{k=1}^{n} x_k < z_2} \prod_{k=1}^{n} f_k(x_k) dx_k}$$

It follows that if we put

$$g_1(x) = f_1(x) \text{ and } g_k(x) = \int_0^x g_{k-1}(x-y) f_k(y) dy \text{ for } k = 2, 3, \ldots, n,$$

we have

$$(6)\qquad P(A_{z_1} \mid A_{z_2}) = \frac{\int_0^{z_1} g_n(x)dx}{\int_0^{z_2} g_n(x)dx} \quad (0<z_1<z_2).$$

Thus the well-known law of composition of probability densities is valid, though some or all of the integrals

$$\int_0^{\infty} f_k(x)dx$$

may be divergent. It follows that the method of Laplace-transforms can be applied also in this case.

The situation is more complex, if it is not supposed that the random variables considered are positive. If ξ and η are two independent, regularly distributed random variables with density functions $f(x)$ and $g(x)$ and we want to determine the density function $h(x)$ of $\xi+\eta$, we can not put in general

$$h(x) = \int_{-\infty}^{+\infty} f(x-y)g(y)\,dy,$$

because this integral will not converge; however if the limit

$$(7)\qquad \lim_{A\to+\infty} \frac{\int_{-A}^{+A} f(x-y)g(y)\,dy}{\int_{-A}^{+A} f(x_0-y)g(y)dy} = h(x)$$

exists, for a fixed x_0, uniformly for all x, then clearly $h(x)$ is the density function of $\xi+\eta$. Thus for instance if η is uniformly distributed on the whole real axis, and $f(x)$ is arbitrary, we have $h(x) = 1$, i. e. $\xi+\eta$ is also uniformly distributed on the real axis.

It is easy to see that if ξ and η are regular random variables on a conditional probability space, and $h(x, y)$ is their joint density function and if

$$(8)\qquad g(y) = \int_{-\infty}^{+\infty} h(x, y)dx$$

is finite, the conditional density function of ξ under the hypothesis $\eta = y$ is

(9) $$f(x \mid y) = \frac{h(x, y)}{g(y)}.$$

Clearly $f(x \mid y)$ is an ordinary density function, i. e.

$$\int_{-\infty}^{+\infty} f(x \mid y)\,dx = 1.$$

If

$$f(x) = \int_{-\infty}^{+\infty} h(x \mid y)\,dy$$

is also finite, it follows from (9), that

(10) $$f(x \mid y) = \frac{g(y \mid x) f(x)}{g(y)},$$

where $g(y \mid x)$ is the conditional density function of η under the condition $\xi = x$. This is the formula of Bayes for conditional probability spaces; it should be noted that $f(x)$ and $g(y)$ must not be ordinary density functions, i. e. we may have

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{+\infty} g(y)\,dy = +\infty.$$

§ 2. A New Deduction of Maxwell's Law

Let us consider an (ideal) gas consisting of N particles with equal masses m. Let ξ_k, η_k, ζ_k denote the components of the velocity of the k-th particle; we suppose that ξ_k, η_k, ζ_k $(k = 1, 2, \dots, N)$ are random variables defined on a conditional probability space $\boldsymbol{P}$. Let us suppose that the conditional probability distribution generated by the variables ξ_k, η_k, ζ_k $(k = 1, 2, \dots, N)$ in the $3N$-dimensional Euclidean space is uniform*, further that this distribution is regular.

It follows that the kinetic energy $\varepsilon_k = \frac{1}{2} m (\xi_k^2 + \eta_k^2 + \zeta_k^2)$ of the k-th molecule has the density function** $\sqrt{x}$ for $0 < x < \infty$ and thus by 2.6. that the kinetic energy $\varepsilon = \sum_{k=1}^{N} \varepsilon_k$ of the whole gas has the density function $x^{\frac{3N}{2}-1}$ $(0 < x < \infty)$.

---

* This supposition is justified by Liouville's theorem mentioned in the introduction.

** Note that if the random variable ξ has the density function x^{α} for $0 < x < \infty$ (α real) and c is a positive constant, $c\xi$ has also the density function x^{α}.

The conditional density function of ε under condition $\varepsilon_k = t$ is clearly $(x-t)^{\frac{3N-5}{2}}$ for $x > t$. It can be easily verified that the conditions for applying formula (10) for $\xi = \varepsilon_k$ and $\eta = \varepsilon$ are fulfilled, and thus we obtain from (10) by some simple calculations that the conditional density function of ε_k with respect to the condition $\varepsilon = E$ is

$$f(x|E) = \frac{2\Gamma\left(\frac{3N}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{3N-3}{2}\right)E}\left(1-\frac{x}{E}\right)^{\frac{3N-5}{2}}\sqrt{\frac{x}{E}}\,; \tag{11}$$

$f(x|E)$ is the density function of the energy ε_k of the k-th particle. Denoting by V_k the velocity of this particle, we have $\varepsilon_k = \frac{1}{2} m V_k^2$, i. e. $V_k = \sqrt{\frac{2\varepsilon_k}{m}}$; thus if $g(x|E)$ denotes the density function of V_k under the hypothesis $\varepsilon = E$, we have $g(x|E) = f\left(\frac{mx^2}{2}\,\middle|\,E\right) mx$ and thus

$$g(x|E) = \sqrt{\frac{2}{\pi}}\left(\frac{m}{E}\right)^{\frac{3}{2}}\frac{\Gamma\left(\frac{3N}{2}\right)}{\Gamma\left(\frac{3N-3}{2}\right)}x^2\left(1-\frac{mx^2}{2E}\right)^{\frac{3N-5}{2}} \tag{12}$$

This is the density function of the velocity of any of the particles. If N is very large, putting $\beta = \frac{3N}{2E}$ and $\sigma = \frac{1}{\sqrt{m\beta}}$, we obtain from (12) by using Stirling's theorem, and the relation $\lim_{n\to\infty}\left(1+\frac{c}{n}\right)^n = e^c$, that approximately

$$g(x|E) \approx \sqrt{\frac{2}{\pi}}\,\frac{x^2}{\sigma^3}\,e^{-\frac{x^2}{2\sigma^2}}. \tag{13}$$

As it is known that $\frac{E}{N} = \frac{3}{2}KT$ where T denotes the absolute temperature of the gas and K the constant of Boltzmann, we have $\beta = \frac{1}{KT}$ and $\sigma = \sqrt{\frac{KT}{m}}$. Thus finally we obtain

$$g(x|E) \approx \sqrt{\frac{2}{\pi}}\left(\frac{m}{KT}\right)^{\frac{3}{2}}x^2 e^{-\frac{mx^2}{2KT}}, \tag{14}$$

i. e. the Maxwell-law of velocity distribution.

Summarising, we may state our result as follows: *if the components of the velocity of each of N particles are inde-*

pendent from each other and from the components of the velocity of other particles, further if all these random variables have regular uniform (a priori) distributions on the whole real axis, then the (a posteriori) probability distribution of the velocity of each particle under the condition that the kinetic energy of the whole system is given, is approximately, if N is large, the Maxwell-distribution (14).

It can also be shown by the same way that the conditional probability function of each of the variables ξ_k, η_k, ζ_k under the condition $\varepsilon = E$ is exactly

$$(15) \qquad h(x\,E) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{m}{E}} \cdot \frac{\Gamma\left(\frac{3N}{2}\right)}{\Gamma\left(\frac{3N-1}{2}\right)} \left(1 - \frac{mx^2}{2E}\right)^{\frac{3N-3}{2}}$$

and thus approximately for large N

$$(16) \qquad h(x\,E) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

where $\sigma = \sqrt{\frac{KT}{m}}$ Thus under the condition $\varepsilon = E$ all these variables are approximately normally distributed with mean value 0 and the same variance σ^2. This is the reason why the velocities have a Maxwell-distribution, because the Maxwell-distribution can be defined as the distribution of a variable $\sqrt{\xi^2+\eta^2+\zeta^2}$, where ξ, η, ζ are independent and normally distributed random variables with mean value 0 and with the same variance σ^2.

If we interprete consequently the „structural functions" of Khintchine as density functions of the distributions of a random variable on a conditional probability space, as defined in 1., we can build up statistical mechanics systematically on a purely probabilistic basis.

Received August 3, 1955

REFERENCES

1. A. Kolmogoroff. Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin, 1933.
2. А. Я. Хинчин. Математические основания статистической механики. Москва Гостехиздат, 1943.
3. А. Я. Хинчин. Математические основания квантовой статистики Москва, Гостехиздат, 1951.
4. A. Rényi. A new axiomatic theory of probability. Acta Math. Acad. Sci. Hung., 6 (1955), p. 185–336.