COMPUTER SIMULATION OF \( n \)-RANDOM VARIABLES WITH ARBITRARY JOINT PROBABILITY DISTRIBUTION THROUGH THEIR REPRESENTATION OVER THE LEBESGUE UNIT CUBE*

VICTOR M. BOGDAN

The Catholic University of America
Washington, DC 20064, USA

Communicated by S. M. Ulam

(Received August 1980)

1. INTRODUCTION

In many problems of quantum chemistry, nuclear and plasma physics, and economics, one meets random processes in which the variables are not independent and which can have a discrete, continuous, or mixed spectrum.

In this paper an algorithm is presented which allows by means of a computer to simulate any such random process \( x_1, \ldots, x_R \) of real random variables if their joint distribution is known. Examples are presented for illustration of the theory.

Mathematical tools which are used in developing the theory are based on Kolmogorov's fundamental paper on probability theory \[1\] and some results of Halmos \[2, 3\], Neveu \[3\], and Bogdanowicz \[4–8\] on measure and integration theory.

Readers who are interested in applications only should concentrate on Sets 1–4 and 8–11 and read the remaining sections as needed to understand the principles. The knowledge of the Lebesgue integral with respect to an abstract measure is essential to understand the proofs. The use of Dirac's delta function is helpful in applied problems.

The theoretical results are formulated in terms of Borel functions, that is the functions measurable with respect to the smallest sigma ring containing all cubes. This class of functions, as was established in Halmos \[2, 3\], coincides for \( R^k \) spaces with the class of Baire functions, that is the smallest class which is closed under sequential limit and contains all continuous functions.

The importance of Baire functions in the general theory of random processes was presented in Ref. 9.

2. COMPUTERIZATION OPERATOR

Let \( F \) be a probability distribution of a real random variable \( x \); i.e., \( F(a) = P\{x < a\} \) for all \( a \in R \). Such a function has the following properties:

(a) \( F \) is nondecreasing on \( R \),

(b) \( F \) is left side continuous on \( R \),

(c) \( F(-\infty) = 0 \) and \( F(\infty) = 1 \).

These properties characterize distributions of real random variables according to Kolmogorov's theorem; i.e., if \( F \) has properties (a), (b), and (c), then there exists a probability space and a random variable \( x \) over it such that \( F \) is its probability distribution.

*This research was done while the author was a National Research Council Senior Resident Research Associate at NASA's Johnson Space Center during his sabbatical leave from the Catholic University of America in 1979.
If a distribution $F$ is absolutely continuous in the Lebesgue sense, then there exists a Lebesgue summable function $f$ on $\mathbb{R}$ such that

$$F(a) = \int_{-\infty}^{a} f(t) \, dt$$

for all $a \in \mathbb{R}$. Such a function $f$ is called a density of the distribution $F$.

In many applications, one meets distributions which do not have Lebesgue summable densities. For example: Let, for $x \in \mathbb{R}$,

$$f(x) = p_1 \delta(x - x_1) + p_2 \delta(x - x_2) + \cdots + p_n \delta(x - x_n),$$

where $0 \leq p_1$ and $p_1 + \cdots + p_n = 1$ and $\delta(x)$ denotes the Dirac's delta function; i.e., the formal density of the distribution $h$ given by the formula

$$h(a) = 0 \quad \text{if} \ a \leq 0,$$

$$h(a) = 1 \quad \text{if} \ a > 0.$$

Then the distribution $F$ corresponding to the function $f$ is given by the formula

$$F(a) = p_1 h(a - x_1) + \cdots + p_n h(a - x_n) \text{ for all } a \in \mathbb{R}.$$

Such a distribution has jumps at the points $x_1, \ldots, x_n$ and is constant between them. Hence $F$ is not an absolutely continuous function. A random variable $x$ corresponding to such a distribution has only discrete states $x_1, \ldots, x_n$. However, in some applications, one meets random variables with discrete and continuous states. For example, consider a random variable with a density $g(x) = p_1 \delta(x - x_1) + p_2 h(x) e^{-x}$, where $0 \leq p_1$ and $p_1 + p_2 = 1$, and $h$ is the distribution of $\delta$. Such a density could appear in a steady-state process $x$ representing the energy level of a particle if the source emits particles having the specific energy level $x = x_1$ with probability $p_1$ and having other energy levels $x \neq x_1$ with joint probability $p_2$.

To stimulate random variables with mixed states, it is convenient to introduce the computerization operator $c$ mapping a distribution $F$ into a function $G$ over the open unit interval $(0,1)$. The function $G$ is defined by the formula $G(u) = \inf\{x \in \mathbb{R} : u < F(x)\}$ for all $u \in (0,1)$.

**Theorem 2.1** The computerization operator $c$ is well defined for every function $F$ being a distribution; i.e., having properties (a), (b), (c). If the variable $u$ has the uniform distribution on the open interval $(0,1)$, then the variable $x = G(u)$ has the probability distribution equal to the function $F$.

**Proof.** Take any $u \in (0,1)$. It follows from property (c) that there are two points $x_1, x_2 \in \mathbb{R}$ such that $F(x_1) < u < F(x_2)$. This implies that the set $A(u) = \{x \in \mathbb{R} : u < F(x)\}$ is nonempty. It follows from property (a) that the number $x_1$ is a lower bound of the set $A(u)$. From the axiom of continuity we get that $G(u) = \inf A(u)$ is well defined.

To compute the distribution of the variable $x = G(u)$, consider the set $H(a) = \{u \in (0,1) : G(u) < a\}$ for a fixed $a \in \mathbb{R}$. It follows from the property of infimum that $u \in H(a)$ if and only if there exists an $x \in \mathbb{R}$ such that $x < a$ and $u < F(x)$. Thus introducing the set $\{u \in (0,1) : u < F(x)\} = (0, F(x))$ we get $H(a) = U_{x < a}(0, F(x))$. Since the function $F$ is continuous from the left, the union of the intervals $(0, F(x))$ over all $x$
< a is equal to the interval \((0, F(a))\). Thus, \(H(a) = (0, F(a))\) for all \(a \in R\). Since the probability that a uniformly distributed variable on the interval \((0,1)\) falls into an interval \(I\) being a subinterval of \((0,1)\) is equal to the length of that interval, we get \(P(x < a) = P(H(a)) = P((0, F(a))) - F(a)\) for all \(a \in R\).

**Remark 2.1.** For every distribution \(F\) the function \(G = c(F)\) has the following properties:

1. \(G\) is nondecreasing on \((0,1)\),
2. \(G\) is right side continuous on \((0,1)\),
3. If \(u \in (F(x), F(x+))\), when \(F\) has a jump at \(x\), then \(G(u) = F(x+)\).
4. If \(F\) is strictly increasing and continuous on a closed interval \((c,d)\) then \(G(u) = x\) for \(u \in (G(c), G(d))\) if and only if \(u = F(x)\).

These properties of the function \(G\) allow one to find the graph of the function \(G\) from the graph of the distribution function \(F\) by the following rules:

1. Fill all vertical jumps in the graph of the function \(u = F(x)\) by linear segments.
2. Treat the \(u\) axis as the axis of the independent variable and the \(x\) axis as the dependent variable.
3. At points \(u\) where there are several values of \(x\) such that \(u = F(x)\) define \(G\) so that \(G(u) = G(u+)\).

### 3. SIMULATION OF A SINGLE RANDOM VARIABLE

Since random number generators available on computers simulate a random variable \(u\) with uniform distribution on the interval \((0,1)\), one may simulate with good accuracy the distribution \(F\) of a random variable whose computerization \(G\) is a piecewise continuous function. Many distributions appearing in applied problems fall into this category.

**Example 3.1.** Assume that each call to the FORTRAN function RAN(0) gives a different random number \(u\) in the interval \((0,1)\). Write a segment of a FORTRAN program to generate \(N = 31\) random values of a variable with the density \(g(x) = 0.56(x + 1) + 0.5h(x)\) \(\in R\), where \(\delta\) and \(h\) are defined as before.

**Solution.** Compute the distribution function from the density \(g\) by the formula

\[
F(x) = \int_{-\infty}^{x} g(t) \, dt
\]

for all \(x \in R\). We get \(F(x) = 0.5h(x + 1) + 0.5h(x)(1 - e^{-x})\). The graph of the function \(F\) is given in Fig. 1.

![Graph of F(x)](image)

From this diagram one gets the formula for the computerization \(G\) of the distribution \(F\) using Remark 2.1. This yields \(G(u) = -1\) if \(0 < u < 0.5\), \(G(u) = -\log(2 - 2u)\) if
0.5 ≤ u < 1. Thus the segment of the program to simulate the distribution \( F \) may look as follows:

```fortran
DIMENSION X(100)
N = 31
DO 5 I = 1, N
X(I) = G(RAN(0))
5 CONTINUE
```

```fortran
FUNCTION G(U)
IF (0.0.LT.U.AND.U.LT.0.5) G = -1.
IF (0.5.LE.U.AND.U.LT.1.) G = -ALOG(2.-2.*U)
RETURN
END
```

The array \( X(I) \) for \( I = 1 \) to 31 will contain a random sample of a variable whose distribution is given by the function \( F \).

**Example 3.2.** Find the computerization \( G \) for a random variable whose density function is \( f(x) = (1/2) \cos x \) if \(-\pi/2 < x < \pi/2\), and \( f(x) = 0 \) if \( |x| \geq \pi/2\).

**Solution.** The distribution \( F \) of the function \( f \) is given by the formula

\[
F(x) = \begin{cases} 
(1/2) (\sin x + 1) & \text{if } |x| \leq \pi/2, \\
0 & \text{if } x < -\pi/2, \\
1 & \text{if } x > \pi/2.
\end{cases}
\]

Since the function \( F \) is continuous on the closed interval \((-\pi/2, \pi/2)\) and maps it onto the interval \((0, 1)\), we can use the equivalence of \( x = G(u) \) with \( u = F(x) \); i.e., \( u = (1/2)(\sin x + 1) \). Solving this equation for \( x \), one gets \( x = \arcsin(2u - 1) \). Thus the computerization \( G \) is given by

\[ G(u) = \arcsin(2u - 1) \]

for \( u \in (0,1) \).

**4. SIMULATION OF INDEPENDENT RANDOM VARIABLES**

In many applications one has to investigate processes consisting of several random variables \( f_1, f_2, \ldots, f_k \). Such variables are called independent if their joint distribution function \( F \) defined by

\[ F(a_1, a_2, \ldots, a_k) = P\{f_1 < a_1, \ldots, f_k < a_k\} \]
for all \((a_1, \ldots, a_k) \in \mathbb{R}^k\), can be represented in the form

\[ F(a_1, a_2, \ldots, a_k) = F_1(a_1)F_2(a_2)\ldots F_k(a_k) \]

for all \((a_1, \ldots, a_k) \in \mathbb{R}^k\).

Thus to simulate such a process it is enough to find the computerizations

\[ G_i = c(F_i) \quad \text{for } i = 1, \ldots, k. \]

Then, if \(u_1, \ldots, u_k\) are \(k\) independent random variables each having uniform distribution on the interval \((0,1)\), the variables

\[ x_1 = G_1(u_1), \ x_2 = G_2(u_2), \ldots, x_k = G_k(u_k) \]

will have the distribution given by the function \(F\).

**Problem.** Let \(p = (\lambda, \phi)\) be a random point on the sphere

\[ S = \{(x, y, z): x^2 + y^2 + z^2 = 1\}, \]

where \(\lambda\) and \(\phi\) are its spherical coordinates. Simulate the uniform distribution on the sphere \(S\).

**Solution.** The probability density \(f\) at the point \(p \in S\) is given by the formula

\[ f(p) = 1/4\pi \quad \text{for } p \in S. \]

Using spherical coordinates one can write a representation of the set \(S\) neglecting sets of measure zero (poles and one meridian) as

\[ S = \{(\lambda, \phi): 0 < \lambda < 2\pi, -\pi/2 < \phi < \pi/2\}. \]

The density function in these coordinates will have the form

\[ f(\lambda, \phi) = f_1(\lambda)f_2(\phi), (\lambda, \phi) \in S, \]

where

\[ f_1(\lambda) = (2\pi)^{-1} \quad \text{for all } \lambda \in (0, 2\pi), \]
\[ f_2(\phi) = 0.5\cos\phi \quad \text{for all } \phi \in (-\pi/2, \pi/2). \]

Thus the joint distribution \(F\) of the variables \(\lambda, \phi\) is given by

\[ F(a_1, a_2) = \int_{-\infty}^{a_1} f_1(\lambda) \, d\lambda \int_{-\infty}^{a_2} f_2(\phi) \, d\phi = F_1(a_1)F_2(a_2) \]

for all \((a_1, a_2) \in \mathbb{R}^2\), where \(F_1\) is the distribution of \(\lambda\) and \(F_2\) the distribution of \(\phi\). This shows that the random variables \(\lambda, \phi\) are independent. Moreover,

\[ F_1(a_1) = a_1/2\pi \quad \text{if } 0 \leq a_1 \leq 2\pi, \]
\[ F_1(a_1) = 0 \quad \text{if } a_1 < 0, \]
\[ F_1(a_1) = 1 \quad \text{if } a_1 > 2\pi. \]
The distribution $F_2$ was discussed in Example 3.2. Thus computerizations of the distributions $F_1$ and $F_2$ are given by

$$G_1(u_1) = 2\pi u_1,$$
$$G_2(u_2) = \arcsin(2u_2 - 1).$$

Hence the variables $\lambda = 2\pi u_1$ and $\phi = \arcsin(2u_2 - 1)$, where $u_1, u_2$ are independent uniform random variables on the interval $(0,1)$ and will simulate a uniform distribution of points on the sphere $S$.

5. INTEGRAL PROPERTY OF THE COMPUTERIZATION OPERATOR

Let $F$ be a distribution of a real random variable; i.e., $F$ satisfies conditions (a), (b), (c) of Sec. 2. Let $G = c(F)$ be the computerization of the distribution $F$. Denote by $V$ the prering (see [4]) consisting of all intervals $I$ of the form $(a,b)$, $(-\infty, a)$, $(a, \infty)$, where $a,b$ are real numbers. Define a set function $\nu$ on $V$ by the formula

$$\nu(a, b) = F(b) - F(a),$$
$$\nu(-\infty, b) = F(b) - F(-\infty) = F(b),$$
$$\nu(a, \infty) = F(\infty) - F(a) = 1 - F(a).$$

One can prove that the set function $\nu$ is countably additive on $V$ and thus forms a volume in the sense of Bogdanowicz [4]. Following the development of the paper [4] denote by $S(V, R)$ the collection of simple functions; i.e., functions of the form

$$s(x) = r_1c_{A_1}(x) + \cdots + r_kc_{A_k}(x) \quad \text{for all } x \in R,$$

where $A_1, \ldots, A_k$ are disjoint sets from the prering $V$ and $r_1, \ldots, r_k$ are real numbers and $c_A$ denotes the characteristic function of the set $A$. The set of simple functions is linear and the following functionals

$$\int s \, d\nu = r_1\nu(A_1) + \cdots + r_k\nu(A_k),$$
$$\| s \| = |r_1|\nu(A_1) + \cdots + |r_k|\nu(A_k),$$

are well defined on it. The first functional is linear and the second forms a seminorm on $S(V, R)$. Moreover,

$$|\int s \, d\nu| \leq \| s \|$$

for all simple functions.

Denote by $N$ the collection of all sets $A$ of $R$ such that for every $\varepsilon > 0$ there exists a countable family $A_t \in V (t \in T)$ such that the set $A$ is contained in the union $\bigcup T A_t$ and $\Sigma_{t \in T} \nu(A_t) < \varepsilon$. Sets of this collection $N$ will be called $\nu$-null sets.

A sequence $s_n \in S(V, R)$ is called basic if there exists a sequence $k_n$ of simple functions and a constant $M$ such that $s_n = k_1 + k_2 + \cdots + k_n$, $\| k_n \| \leq M 4^{-n}$ for all $n$. Denote by $L(\nu, R)$ the set of all functions $f$ for which there exists a basic sequence $s_n$ and a null set $A \in N$ such that the sequence of values $s_n(x)$ converges to the value $f(x)$ if $x \notin A$. Define

$$\| f \| = \lim \| s_n \|, \int f \, d\nu = \lim \int s_n \, d\nu.$$
These are, according to [4], well defined functionals on $L(v, R)$ and the space $L(v, R)$ coincides with the space of Lebesgue summable functions with respect to the Lebesgue measure $\mu$ being the smallest complete measure extending the volume $v$ (see [7]). Moreover, the two integrals

$$\int f \, dv, \int f \, d\mu$$

coincide. In the sequel we shall write

$$\int_a^b f(x) \, dF(x)$$

to denote the integral

$$\int c_{(a,b)} f \, dv.$$

Notice that the classical Lebesgue integral is generated by the function

$$g(u) = u \quad \text{for all } u \in R,$$

which corresponds to the volume

$$v((a, b)) = b - a$$
on the preering $W$ of all bounded right side open intervals.

We shall say that a function $f$ is $v$-summable on a set $A$, or that the integral

$$\int c_A f \, dv$$

exists, if and only if $c_A f \in L(v, R)$.

**Theorem 5.1** Let $F$ be a probability distribution over $R$ and $G$ its computerization. Then, if the right hand integral in the following formula exists in Lebesgue’s sense, then so does the other and they are equal

$$\int f(x) \, dF(x) = \int_0^1 f(G(u)) \, du.$$

**Remark 5.1.** The theorem is valid for the Riemann–Stieltjes integral when the function $F$ is continuous and invertible. Notice that here both functions $F$ and $G$ may have infinite number of discontinuities and neither have to be invertible.

**Remark 5.2.** Let $v$ be the volume generated by the distribution $F$ and $\mu$ the classical Lebesgue measure over the interval $(0,1)$. The above theorem is equivalent to the following. The map $f \rightarrow f \circ G$ imbeds isometrically and isomorphically the Lebesgue space $L(v, R)$ into the space $L(\mu, R)$.

**Proof of Theorem 5.1.** Let $f = c_{(-\infty,a)}$ then $f \circ G = c_{(0,F(a))}$. Indeed $f \circ G(u) = 1$, if and only if, $G(u) \in (a, \infty)$, i.e., $G(u) < a$, which is equivalent, as proved in Sec. 1, to $u \in (0, F(a))$, i.e., to $c_{(0,F(a))}(u) = 1.$
Thus the characteristic function of an interval \((a, b)\) is mapped into the characteristic function of the interval \((F(a), F(b))\). Indeed,

\[
\chi_{(a,b)}\circ G = (\chi_{(-\infty, b)} - \chi_{(-\infty, a)})\circ G = \chi_{(-\infty, a)}\circ G - \chi_{(-\infty, b)}\circ G = \chi_{(0, F(b))} - \chi_{(0, F(a))} = \chi_{(F(a), F(b))}.
\]

Similarly, the characteristic function of \((-\infty, b)\) is mapped into the characteristic function of \((0, F(b))\) and the characteristic function of \((a, \infty)\) into the characteristic function of \((F(a), 1)\).

By the definition of the volume \(v\) on the prering \(V\) and the Lebesgue measure \(\mu\), we get \(v(I) = \mu(G^{-1}(I))\) for every \(I \in V\), since \(c_I \circ G = c_A\) is equivalent to \(A = G^{-1}(I)\).

These observations yield the equality

\[
\int |s| \, dv = \int |s \circ G| \, d\mu \quad \text{for all } s \in S(V, R).
\]

Let \(W\) be the collection of all left side closed subintervals of the interval \((0,1)\). It follows from the definitions of a \(\nu\) null set that if \(A\) is a \(\nu\) null set then \(B = G^{-1}(A)\) has \(\mu\) measure zero. Notice that if \(s_n \in S(V, R)\) is a basic sequence convergent for all \(x \in A\) to the function \(f\) then the sequence \(s_n \circ G\) belongs to the set \(S(W, R)\) and converges for all points \(u \in B\) to the function \(f \circ G\). Thus from the Bogdanowicz definition of the spaces \(L(\nu, R)\) and \(L(\mu, R)\) we get

\[
\int |f| \, dv = \lim \int |s_n| \, dv = \lim \int |s_n \circ G| \, d\mu = \int |f \circ G| \, d\mu
\]

for all functions \(f \in L(\nu, R)\). This proves the theorem.

**Corollary 5.1.** For every bounded Borel function \(f\) and every \(a \in R\) the following equality holds

\[
\int_{-a}^{a} f(x) \, dF(x) = \int_{0}^{F(a)} f(G(u)) \, du.
\]

**Proof.** It follows from the properties of \(\nu\)-measurable functions [5] that every Borel measurable function is \(\nu\)-measurable. Since \(c_{(-\infty, a)}\) is a Borel function, the product \(g = c_{(-\infty, a)} \cdot f\) is also a Borel function. Being \(\nu\)-measurable and bounded by the simple function \(M_{c_{B}} = M_{c_{(-\infty, a)}} + M_{c_{(a, \infty)}}\) for some \(M\), the function \(g\) is \(\nu\)-summable. Since \(g \circ G = c_{(0, F(a))} \cdot f \circ G\), we get from Theorem 5.1 the equality

\[
\int_{-a}^{a} f(x) \, dF(x) = \int_{0}^{F(a)} f(G(u)) \, du.
\]

**6. Existence of Transition Probabilities**

Let \(f_1, f_2, \ldots, f_k\) be real random variables over a probability space. We shall prove that the following conditional distribution

\[
P\{f_1 < a_1, f_2 = a_2, f_3 = a_3, \ldots, f_k = a_k\}
\]

can be well defined as a Borel function of the vector \(a = (a_1, a_2, \ldots, a_k)\) over the space \(R^k\).
Let $F_m$ denote the joint distribution of the variables $f_1, f_2, \ldots, f_m$. Each distribution function $F$ obtained in this way is nondecreasing with respect to the relation $a < b$ on $R^m$ defined to mean $a_i < b_i$ for all $i = 1, \ldots, m$. This means if $a < b$, then $F(a) \leq F(b)$. Moreover, the distribution function is continuous with respect to increasing convergence, i.e., the condition

$$a_j^* \uparrow a_j \quad \text{for } j = 1, \ldots, m$$

implies

$$F(a_1^*, a_2^*, \ldots, a_m^*) \rightarrow F(a_1, a_2, \ldots, a_m).$$

Finally the function $F$ is normalized, i.e.,

$$F(-\infty, \ldots, -\infty) = 0, \quad F(\infty, \ldots, \infty) = 1.$$

According to Kolmogorov's theorem these properties characterize a joint distribution function $F$, i.e., for every such a function there exists a unique Borel probability measure $P$ over $R^m$ such that

$$F(a_1, \ldots, a_m) = P\{e_1 < a_1, \ldots, e_m < a_m\}$$

for all $(a_1, \ldots, a_m) \in R^m$, where $e_j$ are projection functions defined by

$$e_j(a_1, \ldots, a_m) = a_j \quad \text{for all } (a_1, \ldots, a_m) \in R^m.$$

To make the presentation more general it will be convenient to introduce the following notation. If $M$ and $K$ are subsets of the set $\{1, \ldots, k\}$ and $K$ is a proper subset of $M$, then we will write $K < M$. Subsets of this form will be called indexes. The symbol $|K|$ will denote the number of elements of the set $K$.

We will denote by $R^K$ the space of all vectors $x = (x_t)_{t \in K}$, where $x_t \in R$ denotes the component of the vector $x$ with index $t$. If $M$ and $K$ are two disjoint index sets and their union is $S = M \cup K$ then the space $R^S$ can be identified with the product space $R^M \times R^K$ and every vector $x \in R^S$ can be written in the form $x = (x_K, x_M)$ where $x_K \in R^K$ and $x_M \in R^M$.

If $a \in R^T$ we shall denote by $I(a)$ the Cartesian product

$$\times_{t \in T}(\infty, a_t),$$

where $a = (a_t)_{t \in T}$. Sets of this form will be called in the sequel basic cones.

Now if $F_T$ is a probability distribution on $R^T$ and $p_T$ is the corresponding Borel probability over $R^T$ obtained from Kolmogorov's theorem, i.e., $F_T(a) = p_T(I(a))$ for all $a \in R^T$, for disjoint decomposition of the index set $T$ into nonempty sets $S$ and $U$ the following Borel probability $p_S$ is well defined by the formula

$$p_S(A) = p_T(R^U \times A)$$

for all Borel subsets $A$ of the space $R^S$.

This probability in turn generates a probability distribution $F_S$. A function $p_S^x$, will be called a transition probability from probability $p_S$ to probability $p_T$ if its value $p_S^x(A, x)$ is defined for every Borel set $A$ being a subset of the space $R^U$ and every $x \in R^S$, and
moreover the value \( p_s(A, x) \) as a function of the set \( A \) is a probability measure for every \( x \in R^S \), and as a function of the point \( x \) is Borel for every fixed Borel set \( A \), and moreover
\[
p_T(A \times B) = \int_B p_s(A, x) p_S(dx)
\]
for every Borel subset \( A \) of \( R^U \) and every Borel subset \( B \) of \( R^S \) (see [3, p. 73]).

**Theorem 6.1.** For every Borel probability \( p_T \) over \( R^T \), where \( T \) is finite, and every generated Borel probability \( p_S \), where \( S \) is a subset of \( T \) and the difference set \( U = \{ x \in T : x \notin S \} \) is nonempty, there exists a transition probability \( p_T^U \) from the measure \( p_S \) to the measure \( p_T \).

*Proof.* Let \( q_A \), for every fixed Borel set \( A \) contained in \( R^U \), denote the measure, defined by the formula
\[
q_A(B) = p_T(A \times B)
\]
for all Borel sets \( B \) contained in \( R^S \). Since
\[
q_A(B) \leq p_T(R^U \times B) = p_S(B)
\]
for all Borel sets \( B \) in the space \( R^S \), we get from the Randon-Nikodym theorem (see [3,8]) that there exists a Borel function \( f_A \) summable with respect to the measure \( p_S \) such that after a modification on a Borel set \( p_S \)-measure zero we get
\[
0 \leq f_A(x) \leq 1 \quad \text{for all } x \in R^S
\]
and
\[
q_A(B) = \int_B f_A(x) p_S(dx)
\]
for all Borel sets \( B \) contained in the space \( R^S \).

If \( b \in R^U \) let \( I(b) \) denote the Cartesian product \( \times_{t \in U} (-\infty, b_t) \). Every such set \( I(b) \) is a Borel set. A vector \( b \in R^U \) will be called rational if all its components \( b_t \) are rational. We shall also write \( a < b \) for two such vectors if and only if \( a_t < b_t \) for all \( t \in U \).

Let \( h \) be a function given by the formula
\[
h(a, x) = \sup \{ 0, f_{I(b)}(x) : b < a, b \text{ is rational} \}
\]
for all \( a \in R^U \) and all \( x \in R^S \). The set following the supremum operation is nonempty, since it contains zero, and is bounded. Thus the function \( h \) by the axiom of continuity is well defined.

Since both the measure and the integral are continuous under increasing sequential convergence, we get the relation
\[
p_T(I(a) \times B) = \int_B h(a, x) p_S(dx) \quad \text{(A)}
\]
for all \( a \in R^U \) and all Borel sets \( B \) of \( R^U \). Moreover, from the definition of the function \( h \) we get that, for every fixed \( x \), it is nondecreasing, i.e., if for two vectors \( a, c \in R^U \) we have \( a < c \), then \( h(a, x) \leq h(c, x) \). It is also left side continuous, i.e., if \( a^n < a \) for all \( n \)
and the vectors $a^n$ converge increasingly to the vector $a$, then the sequence of values $h(a^n, x)$ converges to the value $h(a, x)$.

Now from relation (A), using the monotone convergence theorem, we get the following relations

$$p_S(B) = \int_B h(\infty, x)\,p_S(dx)$$

and

$$0 = \int_B h(-\infty, x)\,p_S(dx)$$

for all Borel sets $B$ in $\mathbb{R}^S$, where $h(\infty, x)$ denotes the limit in the variable $a$ whose all coordinates tend to $\infty$. The value $h(-\infty, x)$ is understood similarly. Since

$$h(\infty, x) = \lim h(a^n, x) \quad \text{for all } x \in \mathbb{R}^S,$$

where $a^n$ is an increasing sequence of vectors such that each component $a^n_i$ tends to infinity, we get that $h(\infty, x)$ as a function of $x$ is Borel measurable on the space $\mathbb{R}^S$.

Thus from the Radon–Nikodym theorem there exists a set $C$ of $\rho_S$-measure zero such that

$$h(\infty, x) = 1 \text{ and } h(-\infty, x) = 0 \quad \text{if } x \in C.$$ 

Modifying $h$ on this set by putting

$$h(a, x) = g(a) \quad \text{for all } x \in C,$$

where $g$ is any probability distribution on $\mathbb{R}^V$, we get that the value $h(a, x)$ as a function of $x$ is a Borel function for every $a$ and $h(a, x)$ as a function of $a$ is a probability distribution for every fixed $x$. Thus by Kolmogorov's theorem, it generates a unique probability measure $P_S(A, x)$ defined on all Borel sets $A$ of $\mathbb{R}^U$ for every fixed point $x \in \mathbb{R}^S$.

Let us prove that for every fixed Borel set $A$ of $\mathbb{R}^U$ the following two properties hold:

(A) The function $P_S(A, x)$ as a function of $x$ is Borel measurable over the space $\mathbb{R}^S$.

(B) For every Borel set $B$ of $\mathbb{R}^S$, we have

$$P_T(A \times B) = \int_B P_S(A, x)p_S(dx).$$

To this end denote by $M$ the collection of all Borel sets $A$ of $\mathbb{R}^U$ for which properties (A) and (B) hold.

Observe that the sets $I(a)$ belong to $M$. Since $M$ is closed under disjoint finite union, it is also closed under proper differences, i.e., if $A_1$ is a subset of $A_2$ and both sets $A_1$, $A_2$ are in $M$, then also the difference set $A = \{x \in A_1; x \notin A_2\}$ is in $M$. Thus, if $V$ denotes the prering consisting of all intervals of the form

$$(-\infty, a) \quad \text{and} \quad (b, a),$$

where $a$ and $b$ are real numbers, and $V^U$ denotes the prering consisting of all Cartesian products of the form

$$A = \times_{t \in U} A_t,$$
where \( A_t \in V \) for every \( t \in U \), one can prove by induction with respect to the number of bounded intervals \( A_t \) appearing in the representation of the set \( A \), that the prering \( W = V_U \) is contained in the collection \( M \).

Finite disjoint unions of sets from the prering \( W \) form the smallest ring containing \( W \).

Observe that \( M \) is closed under monotone convergence of sets. This implies according to a theorem of Halmos (see [2]) that \( M \) contains the smallest sigma ring generated by \( W \). One can prove that this sigma ring coincides with the sigma ring of all Borel sets of the space \( R^U \). This concludes the proof of the theorem.

Let \( T, S, U \) be as before and \( q_\beta, p_\beta \) be two transition probabilities from probability \( p_S \) to probability \( p_T \).

**Theorem 6.2** There exists a set of \( C \) of \( p_S \)-measure zero such that

\[
p_\beta(A, x) = q_\beta(A, x)
\]

for all Borel sets \( A \) of \( R^U \) and all \( x \notin C \).

**Proof of Theorem 6.2.** It follows from the definition of a transition probability and from the Radon–Nikodym theorem, that for every Borel set \( A \) of \( R^U \) there exists a Borel set \( C(A) \) of \( p_S \)-measure zero such that

\[
p_\beta(A, x) = q_\beta(A, x)
\]

for all \( x \notin C(A) \). Let \( D \) denote the set of all rational points \( b \) of the space \( R^U \) and let \( C \) be the union of the sets \( C(I(b)) \) over all \( b \in D \). Clearly \( C \) is a Borel set of \( p_S \)-measure zero.

Denote by \( M \) the collection of all Borel sets \( A \) of \( R^U \) such that \( p_\beta(A, x) = q_\beta(A, x) \) if \( x \notin C \). This collection contains basic cones \( I(b) \), where \( b \) is a rational vector. It follows from the monotone continuity of a measure that \( M \) contains every set \( I(a) \) for any \( a \in R^U \). The rest of the argument is the same as in Theorem 6.1. This concludes the proof.

Let a function be given \( F_\beta \), defined on the product \( R^U \times R^S \) such that for every fixed \( x \in R^S \) the value \( F_\beta(a, x) \) considered as a function of the variable \( a \) is a probability distribution and for every fixed \( a \in R^U \) considered as a function of \( x \) is a Borel function. Such a function \( F_\beta \) will be called in sequel a transition distribution.

**Theorem 6.3.** There is one-to-one correspondence between transition probabilities \( q_\beta \) and transition distributions \( F_\beta \). This correspondence is given by the relation

\[
q_\beta(I(a), x) = F_\beta(a, x)
\]  

(6.3)

for all \( a \in R^U \) and \( x \in R^S \).

**Proof.** It is clear that every transition probability generates a transition distribution by means of the formula (6.3). To prove that every transition distribution generates a unique transition probability take an arbitrary fixed point \( x \in R^S \) and denote by \( q_\beta(A, x) \) the value of the probability measure defined for a Borel set \( A \) of \( R^U \) and generated from the probability distribution \( F_\beta(a, x) \) by means of the Kolmogorov’s construction.

To prove that \( q_\beta \) is a transition probability it is sufficient to prove that for every fixed Borel set \( A \) the value \( q_\beta(A, x) \) as a function of \( x \in R^S \) is Borel measurable. To this end denote by \( M \) the collection of all Borel sets \( A \) having that property. Notice that all basic
cones \( I(a) \) belong to \( M \). Notice that \( M \) is closed under disjoint finite union and under the monotone convergence of sets. Thus similarly as in the proof of Theorem 6.1 one concludes that \( M \) coincides with the sigma ring of all Borel sets of \( R^U \). This concludes the proof of the theorem.

7. RESOLUTION OF BOREL PROBABILITIES

Let \( T \) be a finite index set and \( U,S \) its disjoint decomposition into nonempty sets. Let \( p_T \) be a Borel probability over \( R^T \) and \( p_S \) the corresponding probability generated over \( R^S \).

The theorems of the previous section show that the measure \( p_T \) generates almost unique representation of \( p_T \) by means of \( q^U \) and \( p_S \) through the formula

\[
p_T(A \times B) = \int_B q^U(A,x)p_S(dx)
\]

for all Borel sets \( A \) of \( R^U \) and \( B \) of \( R^S \). Conversely any pair \( q^U, p_S \) consisting of a transition probability and a probability on Borel sets generates a unique Borel probability over the space \( R^T \) [3, p. 74].

It follows from Kolmogorov’s theorem that condition (A) is equivalent to

\[
p_T(I(a) \times I(b)) = \int_{I(b)} q^U(I(a),x)p_S(dx)
\]

for all \( a \in R^U \) and all \( b \in R^S \).

The necessity of condition (B) is obvious. To prove its sufficiency fix the point \( a \in R^U \) and consider two measures

\[
r_1(B) = p_T(I(a) \times B),
\]

\[
r_2(B) = \int_B q^U(I(a),x)p_S(dx),
\]

for all Borel sets \( B \) of \( R^S \).

Since these measures coincide for every basic cone according to relation (B), they must coincide for \( B = R^S \). If \( r_1(R^S) = 0 \) then both measures are identically zero on all Borel sets \( B \) of \( R^S \). If \( r = r_1(R^S) > 0 \) then dividing the measures \( r_1 \) and \( r_2 \) by the value \( r \) one gets two probability measures which coincide on all basic cones \( I(b) \). Thus these two probability measures have the same distribution function. By Kolmogorov’s uniqueness theorem this implies

\[
r_1(B)/r = r_2(B)/r
\]

for all Borel sets \( B \) of \( R^U \).

Hence,

\[
p_T(I(a) \times B) = \int_B q^U(I(a),x)p_S(dx)
\]

for all Borel sets \( B \) of \( R^U \) and all basic cones \( I(a), a \in R^U \).

Now holding the Borel set \( B \) fixed by a similar argument to the preceding one, we get relation (A) for all Borel sets \( A \) of \( R^U \) and all Borel sets \( B \) of \( R^S \).

Now let us introduce the following notations for points in the spaces \( R^T, R^S \). If
a ∈ R^T then by a_s, where S is a subset of T, we shall denote the point of R^S such that its component having index i ∈ S coincides with the component a_i of a. Thus we have a = a_T. When S = {t} we shall write a_t instead of a_{(t)}. If U and S are disjoint nonempty sets whose union is the set T then the vector a_T can be identified with the pair (a_U, a_S). Notice also that I(a_T) = I(a_U) × I(a_S). Using this convention relation (B) can be written in the equivalent form as

\[ p_T(I(a_T)) = \int_{I(a_U)} q_T(l(a_U), x_S)p_S(dx_S) \]

for all a ∈ R^T.

Thus it will be convenient to introduce a shorthand notation and convention similar to Einstein's convention in tensorial calculus. Namely, the relation given by formula (A) we shall write as

\[ p_T = q_T^S p_S. \]

This will mean that whenever in such a formula a superscript index set S coincides with the subscript index set, we have an integration over a Borel set with respect to the variable x_S.

Notice that the operator \( E \) mapping a pair \( q_T^S, p_S \) into the element \( q_T^S p_S \) preserves convex combinations in both variables \( q_T^S \) and \( p_S \). Thus it is natural to extend it by homogeneity and linearity onto the space \( B_T^S \) of all Borel transition measures (since we will not use this space in the sequel, we leave it to the reader to give precise definition of this space) and onto the space \( B_S \) of all finite Borel measures over the space \( R^S \).

Now let us extend the definition of the transition measure \( p_T^S \) to include the case when either S or U is empty. It is clear that when U = 0 the transition measure \( p_T^S \) is only a point function and the relation

\[ q_S = p_T^S p_S \]

means that \( p_T^S \) is the Radon–Nikodym derivative of the measure \( q_S \) with respect to the measure \( p_S \).

When S = 0 the transition measure \( p_T^S \) does not depend on the point and thus is a probability measure. Thus we assume \( p_T^S = p_U \).

Finally let \( p_\emptyset = 1 \). Then the relation \( q_T = p_T^S p_S \) defines uniquely the element \( q_T \) for any disjoint decomposition \( S, U \) of the index set \( T \).

If the index set \( S \) consists of a single point \( t \) we shall write \( q_T^t \) instead of \( q_T^{(t)} \) and similarly for the set \( U \).

Let \( T = \{1, 2, \ldots, n\} \). Introduce the following notation \( T(j) = \{k: k < j\} \) for \( j = 1, 2, \ldots, n+1 \). Notice that \( T(1) \) denotes the empty set. A sequence of transition-probabilities

\[ q_T^{(j)} \quad \text{for } j = 1, 2, \ldots, n \]

will be called a resolution of the probability \( p_T \) if and only if

\[ p_{T \cup j} = q_T^{(j)} p_{T \cup j} \]

for \( j = 1, 2, \ldots, n \). It follows from the theorems of the previous section that such transition probabilities exist and are almost unique, i.e., \( q_T^{(j)} \) is unique up to a set of \( p_{T \cup j} \) measure zero.
Notice that the value of the transition probability $q_1^T (A, x)$ gives precise meaning to the following conditional probability

$$p_T \{ f_j \in A \mid f_k = a_k \quad \text{for} \quad k = 1, 2, \ldots, j-1 \}$$

if the probability measure $p_T$ is generated by joint distribution of the functions $f_i (t \in T)$. Since by Theorem 6.1, there exists a transition probability $p_{S(j)}$ from the measure $p_1$ to measure $p_{S(j)}$, we can write

$$p_{S(j)} = p_{S(j)}^1 p_1 \quad \text{for} \quad j = 1, \ldots, n,$$

where $S(j) = \{2, 3, \ldots, j-1\}$.

**Theorem 7.1** If a sequence $q_1^T (j = 1, \ldots, n)$ represents a resolution of the probability $p_T$ then there exists a set of $C$ of $p_T$-measure zero such that for every fixed value $x_1 \notin C$ the sequence

$$q_j^T (A, x_{S(j)}, x_1) \quad \text{for} \quad j = 2, \ldots, n$$

as a function of $A$ and the remaining variables $x_t (t \neq 1)$, represents a resolution of the Borel probability $p_S$ defined by the formula

$$p_S (A) = p_{S(j)} (A, x_1)$$

for all Borel sets $A$ of $R^S$, where $S = \{2, 3, \ldots, n\}$.

**Proof.** From the definition of the resolution of the probability $p_T$ we have

$$p_{S(j+1)} = q_j^T p_{S(j)} \quad \text{for} \quad j = 1, 2, \ldots, n.$$  

It follows from the reduction formula for integrals with respect to a probability measure generated by a transition probability [see [3, p. 74] that

$$p_{S(j+1)} (l(a_{S(j+1)}))$$

$$= \int_{l(a_{S(j)})} q^T (l (a_j), x_{S(j)}) p_{S(j)} (dx_{S(j)})$$

$$= \int_{l(a_{S(j)})} \left( \int_{l(a_{S(j+1)})} q^T (l (a_j), x_{S(j)}, x_1) p_{S(j)} (dx_{S(j)}, x_1) \right) p_1 (dx_1) \quad (7.11)$$

$$= \int_{l(a_{S(j+1)})} r^T (l (a_j), x_{S(j+1)}, x_1) p_1 (dx_1),$$

where $r^T$ denotes the transition probability satisfying the condition

$$r^T (l (a_{S(j+1)}), x_1)$$

$$= \int_{l(a_{S(j)})} q^T (l (a_j), x_{S(j)}, x_1) p_{S(j)} (dx_{S(j)}, x_1). \quad (7.12)$$
This expression as a function of $x_1$ according to Neveu [3, p. 74] represents a Borel measurable function and as a function of the variable $a_{S(j+1)}$ represents a probability distribution for every fixed $x_1$. Hence, by Theorem 6.3, it determines a unique transition probability $r_{S(j+1)}$. In this way from formula (7.11) one obtains the representation $p_{T(j+1)} = r_{S(j+1)}p_1$.

It follows from the uniqueness of transition probability from $p_{T(j+1)}$ to $p_T$, Theorem 6.2, that for some set $C$ of $p_T$-measure zero we have

$$p_{S(j+1)}(J(a_{S(j+1)}), x_i) = r_{S(j+1)}(J(a_{S(j+1)}), x_i)$$

for all $a_{S(j+1)} \in R^{S(j+1)}$ and all $x_i \notin C$ and $j = 2, \ldots, n$.

Equalities (7.12) and (7.13) yield

$$p_{S(j+1)}(I(a_i), x_{S(j)}, x_i) = p_{S(j)}(dx_{S(j)}, x_i).$$

For a fixed $x_1 \notin C$ introduce the sequence of probabilities defined by

$$P_{S(j)}(A) = p_{S(j)}(A, x_i)$$

for all Borel sets $A$ of $R^{S(j)}$ and $j = 2, \ldots, n$, and a sequence of transition probabilities defined by the formula

$$Q_{j}^{S(j)}(A, x_{S(j)}) = q_{j}^{S(j)}(A, x_{S(j)}, x_i)$$

for all Borel sets $A$ and $j = 2, \ldots, n$.

One can prove from relation (7.14) that

$$P_{S(j+1)}(A \times R) = P_{S(j)}(A)$$

for all Borel sets $A$ of $R^{S(j)}$ and $j = 2, 3, \ldots, n$. Relations (7.14), (7.15), (7.16), (7.17) prove that the sequence $Q_{j}^{S(j)}$ represents a resolution of the probability $P_{S(j+1)} = p_S$. This completes the proof of the theorem.

8. EXISTENCE OF CONDITIONAL DISTRIBUTION

Let $q_{j}^{S(j)} (j = 1, \ldots, n)$ be a resolution of the probability $p_T$ as defined before. Let $q_{j}^{S(j)}$ denote one of the transition probabilities from this sequence. Define the function $F_{j}^{S(j)}$ by the formula

$$F_{j}^{S(j)}(a_i, a_S) = q_{j}^{S(j)}(I(a_i), a_S)$$

for all $a \in R^U$, where $U$ is the union of the sets $S$ and $\{t\}$.

These functions will be called the *conditional distributions* and the sequence $F_{j}^{S(j)} (j = 1, \ldots, n)$ a *resolution of the distribution* $F_T$.

The conditional distribution $F_{j}^{S(j)}$ gives a precise meaning to the conditional probability

$$p_U \{f_t < a_t \mid f_j = a_j \text{ for } j \in S\} = F_{j}^{S(j)}(a_t, a_S)$$

for all $a \in R^U$, where $p_U$ is the probability generated by the process $f_t (t \in T)$ over $R^U$. 
**Theorem 8.1.** Every conditional distribution $F_{S}^{f}$, where $t \in S$, is a Borel function over the space $R^{U}$.

**Proof.** Let $Z = \{b_1, b_2, \ldots \}$ be the set of all rational points and $Z_{n} = \{b_1, \ldots, b_{n}\}$. For every fixed $b \in Z$ the value $F_{S}^{f}(b, a_{S})$ as a function of $a_{S}$ on $R^{S}$ is Borel as follows from its definition. Define the function $H_{n}$ by the formula

$$H_{n}(a_{t}, a_{S}) = \sup\{0, F_{S}^{f}(b, a_{S}): b \in Z_{n}, b < a_{t}\}$$

for all $a_{t} \in R$ and $a_{S} \in R^{S}$.

These functions are well defined and the domain of the variable $a_{t} \in R$ can be split into finite number of disjoint intervals $I_{j}$ ($j = 0, 1, \ldots, n$) by means of the points of the set $Z_{n}$ so that for each of these intervals the function $H_{n}(a_{t}, a_{S})$ does not depend on $a_{t}$ and is Borel in the variable $a_{S} \in R^{S}$. This follows from the fact that Borel functions are closed under finite supremum operation, and that on the interval $I_{j}$ only a finite number of elements of the set $Z_{n}$ is smaller than $a_{t}$. Thus every function $H_{n}$ can be represented in the form

$$H_{n}(a_{t}, a_{S}) = \sum_{j} c_{j}(a_{t}) G_{j}(a_{S}),$$

where $G_{j}$ are Borel functions on $R^{S}$ and $c_{j}$ is the characteristic function of the interval $I_{j}$. Since the characteristic function of an interval is a Borel function and the composition of a Borel function with a continuous function is a Borel function, we may consider the value

$$c_{j}(a_{t}) = c_{j} \circ e_{t}(a_{t})$$

as a function of $a_{t}$, where $e_{t}$ is the projection function defined by $e_{t}(a_{U}) = a_{t}$ for all $a_{U} \in R^{U}$, as a Borel function over $R^{U}$. Similarly, we may consider the function

$$G_{j}(a_{S}) = G_{j} \circ e_{S}(a_{S})$$

as a Borel function over $R^{U}$.

Since Borel functions are closed under multiplication and addition, the functions $H_{n}$ are Borel over $R^{U}$. Now notice that

$$F_{S}^{f}(a_{t}, a_{S}) = \lim_{n} H_{n}(a_{t}, a_{S}) \quad \text{for all } a \in R^{U}.$$

Thus the function $F_{S}^{f}$ is a Borel function over the space $R^{U}$. This completes the proof.

Notice that every probability distribution $F_{S}$ on $R^{S}$ generates a unique volume $v$ on the product prism $V^{S}$ consisting of all sets of the form

$$A = \times_{t \in S} A_{t},$$

where $A_{t} \in V$ for all $t \in S$, and $V$ consists of all intervals of the form $(-\infty, a)$ and $(a, \infty)$ (see [7]). The Bogdanowicz’s integral with respect to the volume $v$ coincides with the Lebesgue integral generated by the probability measure $p_{S}$.

Thus the integral

$$\int f(x_{S}) v(\text{d}x_{S})$$
is uniquely determined if the distribution function $F_S$ is known. By an integral with respect to the distribution $F_S$ we shall understand

$$\int f(x_S) F_S(dx_S) = \int f(x) \nu(dx).$$

After this definition notice that each function $F^T_{\mathcal{J}}$ from the sequence representing the resolution of the distribution $F_T$ is uniquely determined $F_{T_\mathcal{J}}$ almost everywhere and we have

$$F_{T_{\mathcal{J}+1}}(a_{T_{\mathcal{J}+1}}) = \int_{(a_{T_\mathcal{J}})} F^T_{\mathcal{J}}(a_j, x_{T_\mathcal{J}}) F_{T_\mathcal{J}}(dx_{T_\mathcal{J}})$$

for all $a_{T_{\mathcal{J}+1}} \in R^{T_{\mathcal{J}+1}}, j = 1, \ldots, n$. These formulas can be used to find the resolution. Again we may write for shorthand $F_{T_{\mathcal{J}+1}} = F^T_{\mathcal{J}} F_{T_\mathcal{J}}$ as in the case of transition probabilities.

9. FUNDAMENTAL THEOREM ON SIMULATION

Let $N = \{1, 2, \ldots \}$ and $N(j) = \{k \in N: k < j\}$. Let $F_{N(n+1)}$ be a probability distribution over $R^n$ and let

$$F^n_{j}(j = 1, \ldots, n)$$

be a resolution of the distribution $F_{N(n+1)}$. If $f_1, f_2, \ldots, f_n$ are random variables over some probability space whose joint distribution is $F_{N(n+1)}$ then $F^n_{j}$ is a Borel function making meaningful the following conditional distribution

$$P\{f_j < a_j | f_k = a_k \text{ for all } k \in N(j)\} = F^n_j(a_1, \ldots, a_j)$$

for all $(a_1, \ldots, a_j) \in R^j, j = 1, \ldots, n$.

If $G$ is a function of $j$ variables $a_1, \ldots, a_j$ being a probability distribution with respect to the variable $a_k$ denote by $c_k$ the computerization operator acting onto the $k$th variable, i.e., $H = c_k(G)$ is defined by

$$H(a_1, \ldots, a_k, \ldots, a_j) = \inf \{a_k: u_k < G(a_1, \ldots, a_k, \ldots, a_j)\}$$

for all $a_1, \ldots, a_k, \ldots, a_j \in R$ and $u_k \in (0, 1)$.

The sequence $H_j(j = 1, \ldots, n)$ of functions defined by $H_j = c_j F^n_j$ for $j = 1, \ldots, n$, will be called a computerization of the distribution $F_{N(n+1)}$.

Lemma 9.1. Each function $H_j$ for $j = 1, 2, 3, \ldots, n$ is a Borel function.

Proof. For $j = 1$ the proof is obvious since $H_j$ is monotone. Take any $j > 1$. First let us prove that $H_j$ is Borel in variables $a_1, \ldots, a_{j-1}$.

To this end take any number $a \in R$. Let $G$ denote the function defined by

$$G(a_1, \ldots, a_{j-1}) = F^n_j(a_1, \ldots, a_{j-1}, a)$$

for all $a_1, \ldots, a_{j-1} \in R$. From the equality of the sets

$$\{ (a_1, \ldots, a_{j-1}) : H_j(a_1, \ldots, a_{j-1}, u) < a \} = \{ (a_1, \ldots, a_{j-1}) : u < G(a_1, \ldots, a_{j-1}) \}$$
and the fact that the function $G$ is Borel measurable follows that the function $H_j$ is Borel measurable in the first $j-1$ variables when the $j$th variable $u$ is fixed.

Since in the variable $u$ the function $H_j$ is monotone and right side continuous, we may conclude that $H_j$ is Borel with respect to all its variables jointly similarly as in the proof of Borel measurability of a conditional distribution in Theorem 7.1.

**Theorem 9.2.** Let $H_j$ ($j = 1, \ldots, n$) be a computerization of the distribution $F_{N(n+1)}$. Define recursively the variables

\[
x_1 = H_1(u_1),
\]

\[
x_2 = H_2(x_1, u_2),
\]

\[
x_3 = H_3(x_1, x_2, u_3),
\]

\[
\vdots
\]

\[
x_n = H_n(x_1, \ldots, x_{n-1}, u_n),
\]

where $u_1, u_2, \ldots, u_n$ are independent random variables with uniform distribution over the open interval $(0,1)$. Then the joint probability distribution of the variables $x_1, \ldots, x_n$ coincides with the distribution $F_{N(n+1)}$.

**Proof.** Since each function $H_j$ is Borel, one can prove by induction that each variable $x_j$ as a function of the variables $u_j$ is also Borel. Thus $x_j$ as functions of variables $u_1, u_2, \ldots, u_n$ are Lebesgue measurable. To find their joint distribution we have to compute the Lebesgue measure of the set

\[D(a) = \{(u_1, \ldots, u_n) \in I^n: x_j < u_j \text{ for } j = 1, \ldots, n\},\]

where $I = (0,1)$. From the definition of the variables $x_j$ and the properties of the computerization operator one gets the identity

\[D(a) = \{u \in I^n: u_j < F_{j}^{N(j)}(x_1, \ldots, x_{j-1}, a_j) \text{ for } j = 1, \ldots, n\}.
\]

We will prove the theorem by induction with respect to $n$. For $n = 1$, we have

\[D(a) = \{u \in I: u < F_1^{N(1)}(a)\} = (0, F_1^{N(1)}(a))\]

and the Lebesgue measure of this set is $p(D(a)) = F_1^{N(1)}(a) = F_{N(2)}(a)$ for all $a \in R$.

Assume that the theorem holds for $n = k - 1$. Notice that the set $D(a)$ can be represented in the form

\[D(a_1, \ldots, a_n) = \{u \in I^n: u_j < F_{j}^{N(j)}(a_1), (u_2, \ldots, u_n) \in D_{x_1}(a_2, \ldots, a_n)\},\]

where

\[D_{x_1}(a_2, \ldots, a_n) = \{(u_2, \ldots, u_n) \in I^{n-1}: u_j < F_{j}^{N(j)}(x_1, x_2, \ldots, x_{j-1}, a_j) \text{ for } j = 2, 3, \ldots, n\}.
\]
Notice that for $F$-almost all $x_1 = H_1(a_1)$, the functions $F^{(i)}_{x_1}$ as functions of the remaining variables form a resolution of the distribution $F_{x_1}^{(i+1)}$ in which the value of the first variable is fixed to be $x_1$. From Fubini's theorem one gets

$$p(D(a)) = \int_0^{F_1(a_1)} \left( \int_{a_2, a_3, \ldots, a_n} du_2, du_3, \ldots, du_n \right) du_1$$

$$= \int_0^{F_1(a_1)} F_{2(n+1)}(x_1, a_2, \ldots, a_n) \, du_1.$$

From Corollary 5.1 we get

$$p(D(a)) = \int_{-\infty}^{a_1} F_{2(n+1)}(x_1, a_2, \ldots, a_n) F_x(dx_1).$$

Finally from the properties of a conditional distribution we get $p(D(a)) = F_{2(n+1)}(a_1, \ldots, a_n)$ for all $a \in \mathbb{R}^n$. This proves the theorem.

10. EXAMPLES ON SIMULATION

Let $p$ be the Borel measure generated by means of Kolmogorov's construction over $\mathbb{R}^n$ from a joint probability distribution $F$ of real random variables, $f_1, f_2, \ldots, f_n$.

Such a sequence of random variables we shall call a random process and the smallest closed set $S$ in $\mathbb{R}^n$, whose complement has $p$-measure zero, we shall call the spectrum of the process. That the spectrum is well defined follows from the fact that every open
Computer simulation of \( n \)-random variables

set in \( \mathbb{R}^n \) is the union of a countable family of open spheres having rational centers and rational radii. Thus the union of all open sets of measure zero is a set of measure zero. The complement of that set is the spectrum. Clearly, to define a random process it is sufficient to define the probability measure \( p \) over the spectrum of the process.

**Problem.** Given is a steady flow of elementary particles through a region \( S \) in the form of a unit disk as on the diagram. The intensity of the current \( i \) per unit of area at a point \( (x,y) \) is given by the formula

\[
i(x,y) = 1 + x.
\]

Find the computerization of the process that a particle arrives at the point whose polar coordinates are \( (\lambda, \phi) \).

**Solution.** The spectrum of the process is the closed circle \( S \) (see Fig. 2). The probability density \( f \) that a particle arrives at a point with coordinates \( (x,y) \) is proportional to the intensity of the current \( i \) at the point, i.e., \( f(x,y) = ci(x,y) \) for all \( (x,y) \in S \). This yields the equation

\[
1 = \int_S f(x,y) \, dx \, dy = c \int_S i(x,y) \, dx \, dy = c\pi,
\]

which yields \( c = 1/\pi \).

In the polar coordinates the set \( S \) has a representation

\[
S = \{ (r, \phi) : 0 < r < 1, 0 < \phi < 2\pi \},
\]

neglecting in \( S \) several lines which have measure zero. Since

\[
dx \, dy = r \, dr \, d\phi,
\]

the probability density \( g \) in the polar coordinates is given by

\[
g(r, \phi) = (1/\pi) r(1 + r \cos \phi).
\]

Thus computing the distribution on the spectrum \( S \), we get

\[
F(a_1, a_2) = \left( 1/2\pi \right) a_1^2 + \left( 1/3\pi \right) a_1^3 \sin a_2
\]

for \( 0 < a_1 < 1 \) and \( 0 < a_2 < 2\pi \). This yields the distribution \( F_1^{N(1)}(a_1) = (a_2)^2 \) for all \( a_1 \in (0,1) \). To find the conditional distribution \( F_{1/2} \) notice that the equation

\[
F(a_1, a_2) = \int_{a_1} F_{1/2}^1(r, a_2) F_1^{N(1)}(dr)
\]

is equivalent to

\[
\int_0^{a_1} \int_0^{a_2} g(r, \phi) \, d\phi \, dr = \int_0^{a_1} F_{1/2}(a_1, a_2) 2a_1 \, da_1.
\]
Differentiating this identity with respect to $a_1$, we get the equation

$$
\int_0^{a_2} g(a_1, \phi) \, d\phi = F_2^1(a_1, a_2)2a_1.
$$

Thus,

$$
F_2^1(a_1, a_2) = (1/2) \int_0^{a_2} g(a_1, \phi) \, d\phi = (1/2\pi)(a_2 + a_1 \sin a_2).
$$

To find the computerization of the process $(\lambda, \phi)$ notice that from the continuity of the distribution $F_1$ on the interval $(0,1)$ we get

$$
u_1 = a_1^2, \text{ or } a_1 = (u_1)^{1/2}.
$$

This yields

$$
H_1(u_1) = (u_1)^{1/2} \text{ for all } u_1 \in (0,1).
$$

Similarly by continuity of $F_2^1(a_1, a_2)$ in the second variable $a_2 \in (0,2\pi)$, we get the equation

$$
u_2 = (1/2\pi)(a_2 + a_1 \sin a_2).
$$

This equation with respect to the variable $a_2$ is the Kepler equation. It can be solved either by iteration or by Newton’s method.

Both methods are easily programmable on a computer. Let $a_2 = H_2(a_1, u_2)$ be the solution of the equation as a function of $a_1 \in (0,1)$ and $u_2 \in (0,1)$. The pair $H_1, H_2$ represents a computerization of the process that a particle arrives at the point with coordinates $(\lambda, \phi)$, that is the pair of variables

$$
\lambda = H_1(u_1), \phi = H_2(\lambda, u_2),
$$

where $u_1, u_2$ are independent random variables with uniform distribution over the open unit interval $(0,1)$, will simulate the process.

**Problem.** Assume that it is given a chemical process which generates ions. Consider the following random process $x, y$ where $x$ is the energy level of an ion and $y$ is its life expectancy. Assume that the density of the probability distribution of the process is given by the formula

$$f(x,y) = h(y) e^{x+y} \delta(x + 2) + (x/2)(h(x) - h(x - 1))h(y) e^{-x+y}
$$

for all $(x,y) \in \mathbb{R}^2$. Find a computerization of the process (notice that $\delta$ and $h$ denote here, respectively, the Dirac’s delta function with mass centered at zero and its distribution function over $\mathbb{R}$).

**Solution.** Notice that the spectrum of the process is given in Fig. 3. It consists of an infinite ray at the position $x = -2$ and an infinite strip above the $x$ axis such that $0 \leq x \leq 1$. 
Computing the distribution function $F$ one gets

\[
F(a_1, a_2) = 0 \quad \text{if } a_2 \leq 0 \text{ and } a_1 \in \mathbb{R},
\]

\[
F(a_1, a_2) = 0 \quad \text{if } a_2 > 0 \text{ and } a_1 \leq -2,
\]

\[
F(a_1, a_2) = \frac{1 - e^{-2a_1}}{2} \quad \text{if } a_2 > 0 \text{ and } -2 < a_1 \leq 0,
\]

\[
F(a_1, a_2) = \frac{1}{2}(1 - e^{-2a_2}) + \left(\frac{1}{2}(a_1 + a_2^{-1}(e^{-a_1} - 1))\right)
\]

\[
\text{if } a_2 > 0 \text{ and } 0 < a_1 \leq 1,
\]

\[
F(a_1, a_2) = \frac{1}{2}(1 - e^{-2a_2}) + \left(\frac{1}{2}(1 + a_2^{-1}(e^{-a_2} - 1))\right)
\]

\[
\text{if } a_2 > 0 \text{ and } a_1 > 1.
\]

This yields the formulas

\[
F_{(1)}^{(1)}(a_1, a_2) = 0 \quad \text{if } a_1 \leq -2,
\]

\[
F_{(1)}^{(1)}(a_1, a_2) = \frac{1}{2} \quad \text{if } -2 < a_1 \leq 0,
\]

\[
F_{(1)}^{(1)}(a_1, a_2) = \frac{1}{2}(1 + a_1) \quad \text{if } 0 < a_1 \leq 1,
\]

\[
F_{(1)}^{(1)}(a_1, a_2) = 1 \quad \text{if } a_1 > 1.
\]

The spectrum of the variable $x$ consists of the point $x = -2$ and the interval $(0,1)$. The measure generated by the distribution $F_{(1)}^{(1)}$ has at the point $x = -2$ mass $1/2$ and on the interval $(0,1)$ it has a linear mass density equal to $1/2$. 
Using these properties one gets for the conditional distribution $F_\frac{1}{2}$ the values

$$F_\frac{1}{2}(a_1, a_2) = \begin{cases} 1 - e^{-2a_2} & \text{if } a_1 = -2, \\ 1 - e^{a_1 a_2} & \text{if } 0 < a_1 < 1. \end{cases}$$

Since outside of the spectrum of the variable $x$ the function $F_\frac{1}{2}$ may be defined arbitrarily, set

$$F_\frac{1}{2}(a_1, a_2) = \begin{cases} 1 - e^{-2a_2} & \text{if } a_1 \leq 0, \\ 1 - e^{a_1 a_2} & \text{if } a_1 > 0. \end{cases}$$

From the graph of the function $F_1^{(1)}(u_1)$ we get the formula

$$H_1(u_1) = \begin{cases} -2 & \text{if } 0 < u_1 < 1/2, \\ 2u_1 - 1 & \text{if } 1/2 \leq u_1 < 1. \end{cases}$$

Since the function $F_\frac{1}{2}(a_1, a_2)$ is continuous in the variable $a_2$ for a fixed value of the variable $a_1$ and invertible on the interval $(0, \infty)$, we get

$$H_2(a_1, u_2) = \begin{cases} -(1/2) \log(1 - u_2) & \text{if } a_1 \leq 0, \\ -(1/a_1) \log(1 - u_2) & \text{if } a_1 > 0. \end{cases}$$

The pair of functions $H_1, H_2$ represents a computerization of the process $y, x$. In the above example we considered for the sake of simplicity of the exposition a process having only one spectral line and one continuous areal component. The method used here can be easily extended to the case when the spectrum consists of a sequence of spectral lines and of several two-dimensional components.

11. EXPECTED VALUE OF A FUNCTION OF A PROCESS

Many applications require computation of some parameters of a process, such like covariance matrix, moments, characteristic function, etc. These computations require to find the expectation

$$E(f(x_1, x_2, \ldots, x_k)),$$

where $f$ is a Borel function, which is sufficiently regular, defined on the range (spectrum) of the process $x_1, x_2, \ldots, x_k$.

Computer simulation technique allows one to find the approximate values of the expectation and to establish probabilistic bounds on the error of the expected value. This can be done by involving the central limit theorem if the function $f$ has a finite second moment. Treating the value

$$y = f(x_1, x_2, \ldots, x_k)$$

as a random variable, one may find by simulation of the process sufficiently large stochastically independent samples of the variable $y$. The mean of the sample will approximate the expectation $E(y)$. Since the mean for large samples has approximately normal
distribution, from the sample of the variable $y$ one can easily estimate the variance of the mean, and thus get an idea how accurate is the estimate of the expected value.

12. INTEGRAL FORMULA FOR EXPECTATION

Let $F$ be a probability distribution over $\mathbb{R}^n$ and $H_1, H_2, \ldots, H_n$ its computerization. Define a map $G$ from the cube $I^n$, where $I = (0, 1)$, into $\mathbb{R}^n$ by the formula $x = G(u)$, where

$$
x_1 = H_1(u_1),
$$
$$
x_2 = H_2(x_1, u_2),
$$
$$
x_3 = H_3(x_1, x_2, u_3),
$$
$$
\vdots
$$
$$
x_n = H_n(x_1, \ldots, x_{n-1}, u_n),
$$

for all $u \in I^n$.

Let $\mu$ be the Borel measure corresponding to the probability distribution $F$.

Let $m$ be the classical Lebesgue measure over the cube $I^n$.

**Theorem 12.1.** The map $K$ defined by $K(f) = f \circ G$ establishes linear isometric imbedding of the Lebesgue space $L(\mu, \mathbb{R})$ of summable functions into the Lebesgue space $L(m, \mathbb{R})$.

The proof of the theorem is similar to the corresponding proof for one variable presented in Sec. 4.

**Corollary.** If the right hand side integral in the following formula exists, then so does the other and they are equal,

$$
\int_{\mathbb{R}^n} f(x) \, dF = \int_{I^n} f(G(u)) \, du.
$$

This corollary follows from the above theorem. Definitions of the integral are similar to those of Sec. 4.

13. CONCLUSION

The principle result of this paper is the proof of the existence of a recursive algorithm by means of which one can simulate on the computer any finite sequence $x_1, x_2, \ldots, x_k$ of random variables whose joint distribution $F$ is known. These variables may be dependent and their joint spectrum may have continuous and discrete components.

This result should be useful in applications requiring the Monte Carlo method. In particular in problems of quantum chemistry, nuclear and plasma physics, economics, and stochastic control systems.

A word of caution is appropriate here. Since most computer languages use words of a fixed number of bits to represent numbers, the set of numbers available in such languages is finite. Even if one would use some set of computable real numbers, say all
rational numbers, as one could define for example by means of the PL language of Brainerd-Landweber [10], the set of all numbers available on the computer would be at most countable and thus of Lebesgue measure zero.

Hence it is always possible to find a pathological example of a distribution \( F \) whose computerization \( H \) cannot be simulated by a computer. However in most applications the resulting computerization \( H \) consists of functions which can be well approximated by means of piecewise continuous functions whose computational complexity is not too high.

REFERENCES