Unbounded Probability Theory Compatible with the Probability Theory of Numbers*

V. P. Maslov**

Moscow State University

Received April 3, 2012

Abstract—We study the case in which the values of random variables increase without bound.

DOI: 10.1134/S0001434612050112

Keywords: Erdős theorem for partitio numerorum, independent random variables, conditional probability, Bose distribution, fractional dimension, Bose condensate, dequantization, Lagrangian manifold, tunnel canonical operator.

One of the main problems of modern mathematical statistics is that the “general population” of measured random variables is very large and, although the sample is also large, it is much less than the general population.

In the USSR, the typical sample size in public opinion polls was 3.5 thousand persons.

As an example of the general population, we can take all the citizens of the Russian Federation. We wish to compute their mean income.

There are some quantities fixed by the state, such as the minimum wages and the minimal percent \( \tilde{\mu} \).

The tax \( \tilde{\tau} \) is also determined by the state.

On the other hand, many citizens have zero wages, such as the nonworking family members and the unemployed. For simplicity, these will be called by the general term “unemployed.”

If we construct an order series, “order statistics,” in increasing order of income and partition it into clusters up to 1–3%, then, denoting the values of the random \( \xi \)-income by \( \xi_i \), we obtain the variational series

\[ \left\{ \frac{\xi_i}{T} \right\}_{i=0}^{\infty}, \]

where \( T \) plays the role of the minimal wages established by law. Suppose that \( n \) is the number of trials (sample size) and \( n_i \) is the number of outcomes in which the number \( \frac{\xi_i}{T} \) occurred; then, by the usual definition, the estimate of the sample expectation multiplied by \( n \) trials is

\[ \sum n_i \frac{\xi_i}{T} = M, \quad \sum n_i = n. \quad (1) \]

Thus, the estimate of the sample expectation is

\[ E = \frac{M}{n}. \quad (2) \]

Obviously, if \( \xi_i \) takes unboundedly increasing values, then a random finite sample cannot serve as an even approximate “estimate” of the true expectation.

Therefore, to define the mean per capita income, the general population is divided into parts corresponding to regions representative of particular strata of society for which the conjecture of equilibrium distribution can be used. Sometimes the variation in the income of isolated families is studied (longitudinal study).

*The article was submitted by the author for the English version of the journal.
**E-mail: v.p.maslov@mail.ru
Sometimes the following approach to the measurement of the income distribution of the population is used. It is based on regular undocumented reports of 50,000 persons of which 2000 are Moscovites. The size and structure of this sample is kept unchanged. New persons are included only as replacements for those no longer available. There are practically no rich people, because the participation in such surveys is voluntary. But people below the poverty line may well be represented, because their reports are paid for.

The data obtained is extrapolated by comparing the socio-economic characteristics of the respondents (the region and the type of community, sex, age, household composition, type of work, etc.) with the population census data.

There are no estimates of the accuracy of such a procedure. Only comparisons with “theoretically expected” distributions (of the lognormal type with Pareto tail) are available. It was Pareto who first proposed the power distribution to describe income distributions according to empirical data. Also there are numerous theoretical discussions on whether one can interpret the observed deviations as the presence of many maxima.

As sociologists put it, it is very difficult to predict the unpredictable behavior of a person. If a person is at some point of 3-dimensional space, then this situation is described below for $D = 3$.

However, it is impossible to partition infinite sets into a finite number of finite sets. Therefore, in our case, an estimate of the expectation should be called, more precisely, an empirical expectation.

The quantity

$$\lim_{n \to \infty} \frac{\ln M}{\ln n} = D$$

will be called the *dimension*.

It turns out that, for each $M$ under our assumption (3), there exists a critical number of measurements $N_c \ll M$ such that, as the number of measurements $N > N_c$ increases, our asymptotic accuracy (for example, $O(N^\alpha)$, where $\alpha < 1$) will lead, as a rule, to the presence of unemployed and low-paid workers, which, practically, does not affect the value of $N_c$ within the limits of our accuracy.

**Example.** Consider the random variable taking the integer values

$$i = 1, 2, \ldots, k, \ldots$$

Suppose that $n$ is the number of trials and $n_i, i = 1, \ldots$, are the outcomes corresponding to the values $i$ of the positive integers. Then the empirical expectation is

$$\sum_{i=1}^{\infty} \frac{n_i}{n} = E, \quad \sum_{i=1}^{\infty} \frac{n_i}{n} = 1. \quad (4)$$

The problem is reduced to the number-theoretic problem of finding $n_i$, given the integers $M$ and $n$:

$$\sum_{i=1}^{\infty} in_i = M, \quad \sum_{i=1}^{\infty} n_i = n. \quad (5)$$

What dependence $M(n)$ are we interested in?

To understand this, we must consider the additional example of coin tossing. Suppose that $n$ trials yields sets of sequences of $n_i$ heads and tails. It is well known that most sequences, up to $\sqrt{n}$, will concentrate near the probability $1/2$. This is the simplest example of the concentration of empirical probabilities.

This implies that we must seek a dependence $n(M)$ for which the greatest number of solutions of system (5) occurs. By the way, note that the logarithm to base 2 of the number of solutions is called the *Hartley entropy*. This number was obtained by Erdős. Up to $\sqrt{n}$, it is

$$n = \frac{\sqrt{3/2}}{\pi} \sqrt{M \ln M}.$$
If, in problem (5), the number \( n \) is increased, while \( M \) is kept constant, then the number of solutions will decrease. This implies that if the sums (5) are counted \( \text{off from } 0 \), not from 1, i.e.,

\[
\sum_{i=0}^{\infty} n_i = M, \quad \sum_{i=0}^{\infty} n_i = n,
\]

then the number of solutions does not decrease and will remain constant. Physicists call this property a Bose condensate. They assume that a Bose condensate occurs only in the case where the sequence \( \{i^\alpha\} \) is considered for \( \alpha > 1 \), and that there is no Bose condensate in the case where \( \alpha = 1 \).

In fact, it is the other way around.

According to the Bose–Einstein formulas for \( \alpha = 3/2 \) and \( n \to \infty \), sums of the type

\[
\sum_{i=0}^{\infty} i^{3/2} n_i = M, \quad \sum_{i=0}^{\infty} n_i = n,
\]

are replaced by integrals and it is impossible to isolate the point \( i = 0 \). The condensate phenomenon can occur only in a neighborhood of the point \( i = 0 \), not at the point \( i = 0 \) itself. But, using the author’s discrete parastatistical correction, we see that the condensate appears at the point \( i = 0 \) itself.

In the case \( \alpha \leq 1 \), the Euler–Maclaurin formulas do not apply (i.e., it is impossible to replace these sums by integrals) and, therefore, the point \( i = 0 \) can be isolated.

I will try to explain this Bose condensate effect in a different way. As I have already pointed out (see [1]), the Erdős problem (5), is equivalent to the problem of expansion of a number \( M \) into \( n \) summands. The expansion of \( M \) into one summand has only one variant. The expansion of \( M \) into \( M \) summands has also one variant, namely, the sum of 1’s. Hence, somewhere in the interval, there must be at least one maximum of variants. Erdős calculated it.

For example, the number 5 can be expanded into two summands in two different ways (4+1 and 2+3) (the “partitio numerorum” problem). If we also include 0, then we obtain three variants: 5+0=3+2=4+1. Thus, the inclusion of zero allows us to say that we expand a number into \( k \leq n \) summands. Indeed, the expansion of the number 5 into 3 summands includes all the previous variants: 5+0+0, 3+2+0, 4+1+0 and adds new variants without zero.

Here the maximum is not markedly varied; on the other hand, the number of variants does not change: the zeros, i.e., the Bose condensate, permits the maximum to remain unvaried, and the entropy never decreases, but, reaching the maximum, becomes constant. It is this remarkable property of the entropy that allows us to construct an unbounded probability theory in the general case.

Let us present this construction.

With each dimension we associate a particular basic sequence, which, in thermodynamics, corresponds to a pure gas of identical molecules. Namely, we consider the following basic set of random variables \( \{i^{D/2}\} \) corresponding to the dimension \( D \).

The variational series

\[
\xi_i = i^{D/2},
\]

will be called the basic series of dimension \( D \).

Let \( n_i \) be the value of the outcome corresponding to \( \xi_i \).

The sample value (estimate) of the expectation is

\[
\frac{\sum n_i \xi_i}{n} = E.
\]

Assume that \( E \to \infty \) as \( n \to \infty \).

For convenience, let us also introduce the parameter \( \gamma \) by the formulas \( D = 2(\gamma + 1) \), \( \gamma = D/2 - 1 \). We shall consider the one-parameter family of entropies \( S_\gamma \) as the logarithm of the number of solutions of the system of relations

\[
\sum_{i=1}^{\infty} n_i = n, \quad \sum_{i=1}^{\infty} n_i \xi_i \leq nE.
\]
The asymptotic relation $n(E)$ between $n$ and $E$ as $n \to \infty$ corresponding to the maximal number of solutions of system (8) (up to $o(N)$) is called the maximal accumulation point of the estimate (of the sample value) of the expectation, while the sets of solutions $\{n_i\}/n$ of Eqs. (8) are said to be typical; the value of $n$ at which maximal accumulation is attained is denoted by $N_c$.

The logarithm to base 2 of the number of solutions of relation (8) is the Hartley entropy $S$.

Let $\alpha_1, \alpha_2$ be the Lagrange multipliers for determining the maximum of $S_\gamma$ subject to conditions (8)

$$dS = \alpha_1 \, dn + \alpha_2 \, d(En), \quad (9)$$

$$d(En) = \frac{1}{\alpha_2} \, dS - \frac{\alpha_1}{\alpha_2} \, dn. \quad (10)$$

Denote the quantity $1/\alpha_2$ by $T$ and the quantity $\alpha_1/\alpha_2$ by $\mu$.

The variable $T$ is called the minimal wages and $\mu$ the nominal percentage or the desaccumulation coefficient. We also introduce the notation $b = 1/T$. To the value $n = N_c$ corresponds $T = T_c$.

For a given $E$, the maximum of accumulation is attained at the value of $n(E)$ denoted by $N_c$.

For $\gamma > 0$, the Hartley entropy $S_\gamma$ is of the form

$$S = n \left[ (2 + \gamma) + \frac{\mu}{T} \right]. \quad (11)$$

For $\mu = 0$ and for $n > N_c$, the entropy $S$ is extended continuously by a constant (in the particular case $D = 3$, this phenomenon is called by physicists a “Bose–Einstein condensate”).

For a fractional dimension $D > 2$, passing to integrals by the Euler–Maclaurin formulas, we obtain the one-parameter basic family [2], [3], [4]

$$En = \frac{1}{\Lambda^2(1+\gamma) \Gamma(2+\gamma)} \int_0^\infty t^{1+\gamma} \, dt \left( \frac{1}{\exp(t + \mu/T) - 1} - \frac{n}{\exp n(t + \mu/T) - 1} \right), \quad (12)$$

where $\Lambda = \text{const.}$, as well as

$$n = \frac{1}{\Lambda^2(1+\gamma) \Gamma(2+\gamma)} \int_0^\infty t^\gamma \, dt \left( \frac{1}{\exp(t + \mu/T) - 1} - \frac{n}{\exp n(t + \mu/T) - 1} \right), \quad (13)$$

or, in view of the notation

$$\text{Li}_s(x) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{t^{s-1}}{(e^t/x) - 1}, \quad (14)$$

we have

$$En = \frac{1}{\Lambda^2(1+\gamma) \Gamma(2+\gamma)} \left( \text{Li}_{2+\gamma}(z) - \frac{1}{n^{\gamma+1}} \text{Li}_{2+\gamma}(z^n) \right), \quad (15)$$

where $z = e^{-\mu/T}$, and

$$n = \frac{1}{\Lambda^2(1+\gamma)} \left( \text{Li}_{1+\gamma}(z) - \frac{1}{n^{\gamma}} \text{Li}_{1+\gamma}(z^n) \right). \quad (16)$$

**Remark.** In the case $D \leq 2$ ($\gamma \leq 0$) and also as $\mu \to 0$ ($z \to 1$), the second term in (16) becomes significant.

In [1], the case $\mu = o(1)$, (more precisely, $\mu = o(1/N_c^{\gamma/(\gamma+1)})$) was calculated. Using the author’s discrete parastatistical correction, we obtain

$$N_c = \left( \frac{1}{2} + \frac{1}{3} \frac{\mu}{T} \right) \frac{\Gamma(\gamma + 1)\zeta(\gamma + 1)}{\Lambda^2(\gamma+1)}. \quad (17)$$

---

1 For basic sequences, $\Lambda = 1$ (see [8]).
But if $\tilde{\mu} > \delta$, where $\delta > 0$ and is independent of $N$, then the second integral in formula (16) can be neglected.

For $\mu \equiv 0$ and $\gamma < 0$, the formulas are significantly simplified.

In particular, the following notation was introduced in [2]:

$$F(\xi) = \left(\frac{1}{\xi} - \frac{1}{e^{1/\xi} - 1}\right)$$

and

$$C(\gamma) = \int_0^\infty F(\xi) \xi^\gamma d\xi, \quad -1 < \gamma < 0.$$ 

V. E. Nazaikinskii noted that here the Euler–Maclaurin formulas are not applicable for the passage to integrals and that, using the function $F(\xi)$, one can express the constant in the discrete case quite simply.

Namely, in the discrete case, we have

$$N_c = \sum_{j=1}^\infty \frac{j^\gamma}{ebj - 1} = \sum_{j=1}^\infty j^\gamma \frac{1}{bj} - \sum_{j=1}^\infty j^\gamma F(bj). \quad (17)$$

Since the function $f(x) = x^\gamma F(bx)$ is monotone decreasing, it follows that

$$\sum_{j=1}^\infty j^\gamma F(bj) = \sum_{j=1}^\infty f(j) \leq \int_0^\infty f(x) dx = \int_0^\infty x^\gamma F(bx) dx = b^{\gamma - 1} \int_0^\infty x^\gamma F(x) dx. \quad (18)$$

This inequality will be called the Nazaikinskii inequality.

Thus,

$$N_c = b^{-1}\zeta(1 - \gamma) + O(b^{-1-\gamma}), \quad \gamma < 0, \quad b = \frac{1}{T} \quad (\text{for } \tilde{\mu} = 0, \ T = T_c),$$

where $\zeta$ is the Riemann zeta function.

Thus, in passing from $\tilde{\mu} \equiv 0$ to $\tilde{\mu} = o(1)$, we obtain a jump. And further, for $\tilde{\mu} = \delta > 0$, where $\delta$ is independent of $N$, we obtain another jump in expression (16) without the second term in parentheses.

These jumps correspond to instability jumps when the value of $\tilde{\mu}$ becomes positive under small perturbations.

**Definition.** Events are said to be independent if their respective Hartley entropies add together.

Consider the conditional probability for independent events.

As is well known, by the conditional probability of an event we mean the probability of the simultaneous occurrence of the pair consisting of the event and the condition renormalized by the probability of the condition:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

where $A$ is the event and $B$ is the condition.

The formula

$$P(B) = \sum_{i=1}^k P(B \cap A_i) = \sum_{i=1}^k P(B|A_i)P(A_i), \quad \bigcup_i A_i = \Omega, \quad (19)$$

for disjoint events $\{A_i\}$ is called the formula for total probability.

A similar formula is valid in the new concept of probability. It suffices to obtain it for $k = 2$. First, this is the equality for the entropies taking into account the conditional probability $\alpha$ and the equality $\beta = 1 - \alpha$. Namely, for $\tilde{\mu} = 0$, we have

$$Z_c(\gamma_c + 2) = \alpha(\gamma_1^c + 2)Z_1^c + \beta(\gamma_2^c + 2)Z_2^c. \quad (20)$$
where
\[ Z_c = \frac{\zeta(\gamma + 2)}{\zeta(\gamma + 1)}, \quad \alpha + \beta = 1. \]

Further, obviously, the expectations multiplied by the number of trials for each of them satisfy the relation
\[ (\gamma_c + 1)Z_cT_c = \alpha(\gamma_1^c + 1)Z_1^cT_1^c + \beta(\gamma_2^c + 1)Z_2^cT_2^c. \] (21)

It is of interest to note that, as a result, the highly complicated system of equations of thermodynamics almost coincides with Kay’s empirical rule [5].

Using these formulas, we can find the values of \( T_c \) and \( \gamma_c \) for “mixtures” of the type (19), on the basis of basic variational series.

If we carried out \( N_c \) trials and, further, began to increase the desaccumulation coefficient \( \tilde{\mu} \), then the number of trials cannot decrease, but it remains equal to \( N_c \) (irreversibility of the number of trials), although the estimate of the expectation will vary with the increase of the desaccumulation coefficient.

Consider that \( D > 2 \). Consider 4-dimensional phase space, where for coordinates we take the quantities \( T \) and \( \tilde{\mu} \), and for the corresponding momenta, \( S \) and \( n \). Then, in view of (10), the quantity \( \hat{E}n \) will play the role of the action and determine the two-dimensional Lagrangian manifold \( \lambda_2 \) in 4-dimensional phase space.

Consider the tunnel canonical operator on \( \lambda_2 \) and let the small parameter (diffusion) tend to zero. This procedure is called dequantization [6] and yields a jump in the action \( \hat{E}n \), which, by similarity with its thermodynamic analog, is called the phase transition from an open state to a closed state (this is similar to the transition from the gas state to the liquid state accompanied by a step-wise decrease in entropy).

Let us pass to the Van der Waals normalization, which was also partially used in [7]:
\[ T_{\text{reduced}} \overset{\text{def}}{=} T_r = \frac{T}{\Lambda^2T_c}, \quad [En]_{\text{reduced}} \overset{\text{def}}{=} [En]_r = \frac{En}{E_cN_c}. \] (22)

Then each value of \( T_r \leq 1 \) is associated with the quantity \( N_c = T^{1+\gamma_0}\zeta(1+\gamma_0) \), where \( \gamma_0 \) is determined by the condition
\[ \frac{\zeta(\gamma_0 + 2)}{\zeta(\gamma_0 + 1)} = Z_c = \frac{M_c}{N_cT_c} \]
for \( \tilde{\mu} = 0 \).

In view of (10) and the equality \( dn = 0 \), for a closed system in which \( n \geq N_c \) satisfies the relation \( n \equiv N_c \) (an analog of the Bose condensate), the maximum of the entropy with respect to \( \tilde{\mu} \) equal to \( S = T_r^{\gamma+1}\frac{\zeta(\gamma + 2)}{\zeta(\gamma_0 + 2)} \)
is attained with respect to \( \gamma \) at \( \gamma = 0 \). Here
\[ \gamma = \gamma(Z), \quad \text{where} \quad Z = \frac{\zeta(\gamma + 2)}{\zeta(\gamma + 1)}. \]

Hence
\[ Z = \frac{\text{Li}_2(e^{\beta/T_r})}{T_r^{\gamma_0+2}\zeta(\gamma_0 + 1)}. \]

In the case \( Z = Z_c \), the final point is obtained for \( M_c = T_r^{\gamma_0+2} \). Using the condition \( Z = Z_c \), we find the value of \( \tilde{\mu} \) at the final point.

On the plane \( Z, M \) this is the straight line passing from the point \( Z = 0, M = 0 \) to the final point indicated above.
The equality of the actions under dequantization implies the condition of the equality of \( \tilde{\mu} \) and \( \tilde{\mu} \), which determines the phase transition from an open system (similar to the gas state) to a closed system (of the liquid type); this condition coincides with Maxwell’s rule in thermodynamics.

The ideas of this paper stemmed from the conjectures put forward by T. V. Maslova in her dissertation for the Master’s degree. The author used an approach proposed by T. V. Maslova for testing context–dependent applications, in particular, for enumerating the sample values of the probabilities of requirements for carrying out tests. When the text of the program being tested is unknown, then we can apply the conception given here. Taking context–dependent tests into account, we obtain some auxiliary constraints, which increases the dimension of the phase space.

Conclusions for Incomes

First, the following question arises: How to count money? In each currency, there is a minimal unit (copeck, cent, penny, etc.). Namely, these units are used to count, in particular, the minimal wages per year. Significantly, this is a very large number. In the proposed conception of the unbounded theory of probability, the principal aspect is that all the numbers tend to infinity.

The incomes in European countries can be classified according to the progressive tax scale. In England, the taxes do not amount to more than 50% of the income. In Netherlands, they exceed 50% of the income, and since the percentage of people evading the taxes is small, the order scale of incomes can be determined. If this scale increases faster than the linear law: \( \xi_i \geq i^\alpha \), where \( \alpha > 1 \), then phase transitions from the open state to the closed state may occur, i.e., from the state in which the income in a given currency can be freely converted into a European or American currency to the state in which such a conversion is forbidden. Such closed systems could be found, for example, in England (after the second world war), in Israel, and in the USSR.

The next point of interest is the expansion of the variational series \( \{\xi_i\} \) increasing faster than \( i^\alpha \) in terms of the basic series by formulas (20)–(21). This can be done in some approximation and is even rigorously justified in the framework of nonstandard analysis (E. V. Shchepin’s oral communication).

The following conclusions can be drawn. There exist critical wages such that, above them, phase transitions do not occur. If the ratio of the minimal wages to the critical wages is known, then one can indicate the value of the nominal percentage for which the transition from an open system to a closed one occurs.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (grant no. 11-01-12058_oîî_m).

REFERENCES