# NORMAL SEQUENCES

By S. Ducray

*O.* THE PROBLEM. The word 'normal' is used here as in a classical theorem of E. Borel: *Almost all numbers are normal in a 'decimal' system to any base.* That is, given any integer $m \geqslant 2$ with the 'digits' $0, 1, 2, \ldots, m-1$ as the base. Then a decimal to the base $m$ is $a_1 a_2 a_3 \ldots = a_1/m + a_2/m^2 + a_3/m^3 \ldots$, where the $a$'s form an infinite sequence of the given digits. To avoid duplication, the convention is made that a number may terminate in an infinite sequence of 0's, but not in an infinite sequence of the greatest digit, $m-1$. Then there is a $1-1$ correspondence between the set of all decimals and the points of the line segment $0 \leqslant t < 1$. 'Almost all' means except for a set of Lebesgue outer measure zero on $(0, 1]$. A number is 'normal' if any finite combination of $k$ preassigned digits occurs with its proper frequency, *i.e*, as for any other combination of the same number of digits, namely $1/m^k$.

Though this is a theorem in number theory[1], standard books on advanced probability theory[2] would treat it as a special case of the law of large numbers. The latter treatment enables the further question raised by Hausdorff, Hardy, Littlewood and others, to be answered: To what extent may the general number in the system deviate from normality? This is done by the law of large numbers (*LLN*) the central limit theorem (*CLT*), and the law of the iterated logarithm (*LIL*), which has two cases, the upper (*ULIL*) and the lower (*LLIL*). We shall use these abbreviations and the code letter *F* with page number for citing the probability text[2] referred to above. The problem is now to see the content of the theorem of Borel, and of the questions that ensue as regards deviation from normality.

*1.* THE CANONICAL MAPPING. The following standard definitions will be used: A *proper frequency distribution* is furnished by a denumerable set of real numbers $f_i < 0$ such that $\Sigma f_i = 1$. If $A_1, A_2, A_3, \ldots$ be an indexed set of distinct *attributes*, an infinite sequence thereof such as $A_i A_j A_k \ldots$ (not all distinct) constitutes a *sample* or *point of sample-space.* Any sequence $\{ A_r \}$ wherein the limiting frequency with which each particular $A_i$ occurs is the $f_i$ above has that *basic* distribution. By *probability* is meant a measure, obeying the usual postulates, defined over the whole or over a sub-set of sample-space, such that the total measure of the universe of definition is unity. The probability measure of an *event* ( = sub-set of sample-space) is indicated by the

Reprinted from the Journal of the University of Bombay, Volume XXXI, Parts 3 & 5, November 1962-March 1963.

letter $P$. The $n$-th term of a generic sequence is denoted by $X_n$ and $P(X_i = A_j)$ is the probability measure, if it exists, of the set of sample-sequences over which $A_j$ appears as the $n$-th term. The *joint* probability of a *compound* event is similarly defined ; *e.g,* : $P(X_i = A_k\,;\, X_j = A_r, \ldots)$. A sample-sequence is *normal* if every finite combination $A_i A_j A_k \ldots$ occurs with frequency equal to the product of the individual component frequencies $f_i f_j f_k \ldots$ It is understood that normality is defined only when the attributes have a basic, proper frequency distribution, a frequency being the limit of the ratio : The number of times a given $A$ occurs in the first $N$ terms of the sequence, divided by $N$, as $N \to \infty$. Corresponding to normality, the events $X_i = A_k, X_j = A_r, \ldots$ are *independent in probability* if for every finite number of such events the compound probability is the product of the component individual probabilities. The following proposition holds:.

*Theorem* 1 : *Given a set of distinct attributes* $A_0, A_1, A_2, \ldots$ *and a corresponding proper frequency distribution. Then there exists a mapping whereby* : 1) *the totality of sample-sequences is mapped in a* $1-1$ *manner onto* $0 \leqslant t < 1$. 2) *The Lebesgue outer measure on the map is equivalent to probability measure over the sample-space.* 3) *Almost all sample-sequences are normal with respect to the given basic frequency distribution.* 4) *The events* $X_i = A_j$ *are all independent in probability.*

*Proof*: It suffices to construct the actual mapping, as follows. Divide $(0, 1]$ into right-open sub-intervals by marking off successive points : $t_0 = f_0, t_1 = f_0 + f_1, \ldots t_i = f_i + t_{i-1} \ldots$ . Then subdivide the sub-intervals $(0, t_0], (t_0, t_1], \ldots$ in the same manner but *each in proportion to its own length*. And so on, step by step. For mapping a given sequence $A_i A_j A_k \ldots$ , take first the sub-interval immediately to the *left* of $t_i$ in the first sub-division. Then, in the next subdivsion, the sub-interval to the left of the point marked with subscript $j$, within the first sub-interval chosen. And so on, taking the next stage of subdivision for each successive subscript of the sequance. The sequence of nesting intervals obviously converges to a single point in $(0, 1]$, where a suitable convention has to be made (as in the opening section) to exclude sequences ending in infinitely many repetitions of the final attribute if the number of attributes be finite. Conversely, to each such point, there corresponds just one sequence of subscripts (with the convention made for the finite case). The properties listed follow obviously.

The theorem of Borel is a special case when there are only a finite number of attributes, all with equal frequencies. The attribute is there assigned the numerical value of its subscript. It has been shown here that the theorem does not depend in essence, upon any number system, nor upon a finite number of frequencies nor upon equal frequencies.

2. THE ITERATED LOGARITHM LAWS. Hereafter, assign the numerical value $a_i$ to the attribute $A_i$ all the $a$'s being distinct from each other. The $X_n$ all become stochastic variables with the identical probability distribution, defined by $a_i$ and $f_i$. The expectation $E(g(X))$ of any function of the stochastic variable is defined as $\Sigma\, g(a_r) f_r$, provided the sum exists. The standard notation is used :

$$(1)\ldots E(X) = \mu\,;\ E(X^2) - \mu^2 = \sigma^2\,;\ \phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

These are the mean, variance and Gaussian distribution respectively.

Three corollaries may now be asserted for theorem 1, taking $(S_n = X_1 + \ldots X_n)$.

*Cor.*1 : *If* $E(X)$ *exists, then with unit probability,* $\lim S_n/n = \mu$ *as* $n \to \infty$ .

This is Khinchin's form of *LLN* (*F* 191) and is an obvious consequence of the canonical mapping of theorem 1, with our definitions.

*Cor.* 2. *If* $E(X^2)$ *exists, then with unit probability, as* $n \to \infty$ *we have*: $(2)\ldots P(\alpha\sigma\sqrt{n} < S_n - n\mu < \beta\,\sigma\sqrt{n}) \to \phi(\beta) - \phi(\alpha)$.

This is the Lindeberg *CLT* (*F* 192). The proof, a bit more complicated than for the preceding, is quite well known.[3]

*Cor.* 3 : *If* $E(X^b)$ *exists for some* $b > 2$, *then with unit probability, each of the two inequalities*

$(3)\ldots -(1+\varepsilon)\sigma\sqrt{2n \log\log(n\sigma^2)} < S_n - n\mu < (1+\varepsilon)\sigma\sqrt{2n\log\log(n\sigma^2)}$ *holds for every* $\varepsilon > 0$ *with at most a finite number of exceptions as* $n \to \infty$. *But if* $\varepsilon$ *be replaced by* $-\varepsilon$, *then each of the inequalities in* (3) *is false infinitely often.*

The first of these statements is *ULIL* and the second *LLIL*. The proof is classical[4] with considerable refinement possible for the gap between the two laws here given as for $\pm\,\varepsilon$.

It has already been noted that Cor. 1 can be restated as for the theorem of Borel without any number-system at all. Moreover, none of these corollaries need be considered over the whole sample space of *all* possible sequences. For example, taking decimals to the base *3*, omit all points of (0, 1) whose coordinate in decimal expansion involves the digit 1. This is the Cantor ternary set (discontinuum), of Lebesgue measure 0 on (0, 1)—but not denumerable. Nevertheless, the corollaries do hold for almost all sequences of the set, with suitable change of the constants involved in the limit. That is, the frequency distribution is no longer *proper*, and we have to remap upon (0, 1] using the binary (two-attribute) system, in which all three corollaries hold quite obviously. In other words, it is the measure induced *on* the set which matters rather than the Lebesgue measure *of* the the set in the canonical mapping of the previous section. Any projection with induced probability measure will serve, provided it is not entirely a projection of points of some exceptional set in the canonical mapping. The question of independence is bypassed for *ULIL* as follows :

*Theorem 2 : Given a set of sample-sequences, all distinct, in* $1-1$ *correspondence with the points of* (0, 1] *such that Lebesgue measure on the unit interval is equivalent to probability measure over the set; and further :* 1) *All* $X_r$ *have the identical basic (probability and frequency) distribution.* 2) $E(|X|^b)x$ *exists for some* $b > 2$. 3) *The joint distribution of any n consecutive X's is such that CLT holds and* 4) *there exists a constant* $\lambda > 1$ *such that the probability for* $|S_k - k\mu|$ *exceeding or equalling* $\lambda\sigma\sqrt{2n \log\log(n\sigma^2)}$ *for at least one* $k \leqslant n$ *is not greater than in the case of independence of* the *X's. Then ULIL holds as in* (3) *for every* $\varepsilon > \lambda - 1$, *with unit probability measure.*

*Proof* : It suffices here only to note that the *ULIL* is based upon the *first* Borel-Cantelli lemma (*F* 154), *which does not require independence in probability.* The standard proof of *ULIL* (*F* 157—9 for binomial distributions, easily generalized) then depends only upon the conditions given above, and can be carried through step by step . Of course, *LLIL* cannot be treated in this manner with suitably changed inequalites, because it rests upon the *second* Borel-Cantelli lemma (*F* 155), which does require independence in probability.

This paper would not have been possible without the constant labour of Prof. D. D. Kosambi.

## References

(1) G. H. Hardy and E. M. Wright : *An Introduction To The Theory Of Numbers*, 2nd, ed. (Oxford, 1945) ; sections 9.12 and 9.13.

(2) W. Feller : *Introduction To Probability Theory and Its Applications*, vol. I, (New York, 1950) ; pp. 161-163.

(3) J. W. Lindeberg, *Math Zeit.* 15 1922. 211-225; but J. V. Uspensky in his *Introduction to Mathematical Probability* credits Lyapunov with the essential device.

(4) W. Feller : The general form of the so-called law of the iterated logarithm; *Trans. Am. Math. Soc.* 54. 1943. 373-402.