BOLYAI SOCIETY
MATHEMATICAL STUDIES, 14

A Panorama of Hungarian Mathematics in the Twentieth Century, pp. 523–535.

Stochastics: Information Theory

IMRE CSISZÁR

Information Theory has been created by Claude Shannon as a mathematical theory of communication. His fundamental paper {19} appeared in 1948. This was one of the major discoveries of the 20th century, establishing theoretical foundations for communication engineering and information technology. The key ingredients of Shannon's work were (i) a stochastic model of communication, (ii) the view of information as a commodity whose amount can be measured without regard to meaning, and (iii) the emphasis of coding as a means to enhance information storage and transmission, in particular, to achieve reliable transmission over unreliable channels.

Today, the mathematical discipline built on these ideas is often called Shannon theory, while the term information theory is frequently used in a much broader sense. In the terminology we use here, information theory – abbreviated as IT – is a branch of stochastic mathematics whose characteristic tools are mathematical expressions interpreted as measures of information. Problems relevant to information transmission and storage that involve coding represent a central but by no means the only subject of this theory. In fact, IT ideas turned out to be very useful in various fields of pure and applied mathematics, such as combinatorial analysis, ergodic theory, mathematical statistics, probability theory, etc. (But statistical investigations using the measure of statistical information introduced by Ronald A. Fisher in 1925 are not considered pertaining to IT.)

Though IT was created effectively by Shannon alone, ideas of other scientists did influence its birth and early development. Among these scientists there were several Hungarians.

Shannon's entropy, the basic measure of the amount of information, has a close relationship to the concept of entropy in physics. The first to relate information and (physical) entropy was the Hungarian physicist *Leo*

Szilárd (Z. Physik, Vol. 53, 1929, p. 840). A key inequality of IT known as Kraft's inequality is sometimes attributed also to Szilárd (e.g. in [154]), though I could not confirm the correctness of this attribution. Of course, Shannon did rely on the theory of communication available at the time, e.g., he gave special credit to Norbert Wiener for the "formulation of communication theory as a statistical problem". A famous Hungarian contributor to communication theory, explicitly mentioned by Shannon in {20} (on page 11) was *Dénes Gábor* (Nobel laureate, inventor of holography).

As a forerunner of IT, the work of *Ábrahám Wald* on sequential analysis [192] deserves special emphasis, although the IT aspect of this outstanding contribution to statistics has been recognized only later. The IT approach to statistics is generally associated with the name of Solomon Kullback whose book [98] systematically develops statistical applications of the information-theoretic measure of distance of probability distributions, now called information divergence (I-divergence), or Kullback–Leibler distance (also known as relative entropy or information gain). It was, however, Wald who first made essential use of I-divergence, without giving it a name. Kullback (loc. cit., p. 2) writes: "Although Wald did not explicitly mention information in his treatment of sequential analysis, it should be noted that his work must be considered a major contribution to the statistical applications of information theory." For further details on Wald's work in this regard we refer to the Section on Mathematical Statistics.

Information theory in Hungary

Within Hungary, research in IT was initiated by *Alfréd Rényi,* in the fifties (but the first to write about IT in Hungary was *Albert Korodi,* electrical engineer, former coworker of Szilárd). Rényi wrote cca. 25 research papers on IT and, not less importantly, started teaching IT at the Loránd Eötvös University, Budapest; his Probability Theory textbook [152] includes an Appendix on IT. The Eötvös University is still rather exceptional in having IT in the curriculum of mathematics students; elsewhere, IT is mostly pursued in electrical engineering departments. Among the mathematicians covered in this volume, in addition to Rényi it was *István Vincze* who devoted several papers to IT; these mainly address statistical applications of IT. The extraordinarily rich life-work of *Paul Erdős* also contains some papers related to IT, though this certainly was a side-issue for him. Today, the IT research group in Budapest established by Rényi enjoys international reputation. Its contributions are outside the scope of this volume, but some of them directly continuing and complementing Rényi's work will be briefly mentioned below.

Rényi always preferred brand new problems to already much investigated ones, and also in IT he was looking for new directions as opposed to the mainstream subject of coding theorems. His works on IT were mostly concentrated around the following subjects: (i) amount of information for non-discrete distributions ("dimensional entropy") (ii) information measures different from Shannon's ("Rényi informations") (iii) axiomatic characterization of information measures (iv) asymptotic evaluation of the amount of information provided by a statistical experiment. Rényi also initiated a systematic development of search theory that he regarded as a part of IT.

Dimensional entropy

Rényi's first paper on IT, joint with *János Balatoni,* appeared in 1956, see the Selected Papers [151], paper [151, article 121]; also in the sequel, Rényi's papers will be referred to by their number in [151]. The main contribution of the paper [151, article 121] was to clarify the relationship of Shannon's entropy formulas for discrete and continuous distributions, via the concept of "dimensional entropy." Rényi returned to this subject in several subsequent papers, in particular, the results of the first publication were substantially strengthened three years later [151, article 160]; they are reviewed below as appearing there.

The entropy of a discrete random variable ξ whose distribution is $P = (p_1, p_2, \ldots)$ is defined as

$$H(\xi) = H(P) = -\sum p_k \log_2 p_k,$$

and the entropy of a real-valued or vector-valued random variable with density $f(x)$ is defined as $-\int f(x) \log_2 f(x)\, dx$. While Shannon's results convincingly support the interpretation of discrete entropy as a measure of information content, this interpretation does not directly carry over to continuous entropy. Rényi gave a precise meaning to the interpretation of continuous entropy as a "measure of information content up to an infinitely large additive constant."

Rényi defined the (information theoretic) dimension and the d-dimensional entropy of a real-valued random variable ξ via the discrete approximations $\xi^{(n)} = [n\xi]/n$, $n = 1, 2, \dots$ of ξ, by

$$d(\xi) = \lim_{n\to\infty} \frac{H\big(\xi^{(n)}\big)}{\log_2 n}$$

$$H_d(\xi) = \lim_{n\to\infty} \Big(H\big(\xi^{(n)}\big) - d\log_2 n\Big)$$

(the dimension resp. d-dimensional entropy of ξ is undefined if the corresponding limit does not exist). He proved the following: Suppose that $\xi^{(1)} = [\xi]$ has finite entropy. Then, if ξ is discrete, $d(\xi) = 0$ and $H_0(\xi) = H(\xi)$, while if ξ has a density then $d(\xi) = 1$ and $H_1(\xi)$ is given by Shannon's integral entropy formula. Moreover, if the distribution of ξ is a mixture of a discrete and a continuous component, the latter having a density, then ξ has dimension equal to the weight of the continuous component; the corresponding d-dimensional entropy was also determined. Extensions of these results to vector-valued random variables were also treated; the information theoretic dimension of an $\mathbb{R}^d$-valued ξ having a density equals the geometric dimension d, and the corresponding d-dimensional entropy is given by Shannon's integral formula. On the other hand, even for ξ taking values in the unit interval, the dimension need not exist if ξ has a singular distribution.

Rényi [151, article 160] also considered discrete approximations other than $\xi^{(n)}$ above. To any partition π of the range X of ξ into disjoint subsets X_k there corresponds an approximation ξ_π defined by $\xi_\pi = k$ if $\xi \in X_k$. Extensions of "dimensional entropy" results to approximations of this kind were given for the case when X was the unit interval and the X_k's were intervals of length $\le \varepsilon$ with $\varepsilon \to 0$. At the same time *Vincze* (Matematikai Lapok, vol. 10, 1959, pp. 255–266) considered approximations of a real-valued ξ corresponding to partitions into intervals "of equal interest," namely to partitions $\pi_{n,\Phi}$ of the real line into intervals of Φ-measure $\frac{1}{n}$, where Φ is a given probability measure interpreted as the distribution of our interest. He showed (assuming regularity conditions) that

$$\lim_{n\to\infty} \Big(H\big(\xi_{\pi_{n,\Phi}}\big) - \log_2 n\Big) = -\int \log_2 \frac{dF}{d\Phi}\, dF,$$

where F is the distribution of ξ. The integral here is the I-divergence of F from Φ, thus Vincze's result provided an interesting new interpretation of I-divergence, as a measure of information relative to the distribution of our interest.

The concept of dimensional entropy was extended to a class of stochastic processes by M. Rudemo {18}. Using this extension, Rényi immediately established a maximum entropy property of Poisson processes, see [151, article 228]: The maximal dimension of a homogeneous point process in $(0, T)$ with density λ is equal to λT, and the Poisson process has the largest λT dimensional entropy.

Let us mention some later developments, complementing Rényi's work on dimensional entropy. *Imre Csiszár* in {4} proved the following: For a measure space (X, μ) where μ is a non-atomic σ-finite measure on X, consider partitions π_ε of X into subsets of equal μ-measure ε, and let ξ be an X-valued random variable. Then, subject to mild regularity conditions, $H\big(\xi_{\pi_\varepsilon}\big) - \log_2 \frac{1}{\varepsilon}$ converges as $\varepsilon \to 0$ to the generalized entropy of ξ with respect to μ, that equals standard (continuous) entropy when $X = \mathbb{R}^d$ and μ is the Lebesque measure, and equals negative I-divergence when μ is a probability measure. *József Fritz* {11} proved the extension of Rényi's Poisson process result to point processes on arbitrary (nonatomic, separable) measure spaces. Rényi did point out a relationship of his dimension (of random variables) to Hausdorff dimension (of sets), see [151, article 175]; *Péter Gács* in {12} suggested a modified definition of information theoretic dimension that leads to an even closer relationship of this kind.

Rényi informations

Rényi's most widely known contribution to IT was to show that certain quantities different from Shannon's information measures come also into account as alternatives to the latter. These "informations of order α" are now called Rényi informations. Rényi's first publication about them appeared in 1960, where he noted that "information quantities of order α were already investigated in the literature from other viewpoints," the new results consisted in "showing that some reasonable postulates can be satisfied only by them and by Shannon's entroy, and ... how the known results on Shannon's entropy generalize to information measures of order α." Rényi's best known work on this subject is his Berkeley Symposium contribution [151, article 180], one of Rényi's most often cited papers.

Shannon's entropy $H(P)$ of a probability distribution $P = (p_1, \dots p_n)$, measuring the average amount of information provided by a random experiment whose possible outcomes have probabilities $p_1, \dots, p_n$, equals the (weighted) arithmetic mean of the individual informations $I_k = \log_2 \frac{1}{p_k}$ associated with these outcomes.

The arithmetic mean is a special case (φ linear) of means of form $\varphi^{-1}\big(\sum p_k \varphi(I_k)\big)$, where φ is some strictly monotonic function. Rényi argued that means with non-linear φ might also be used, provided they satisfy the intuitive requirement of additivity for independent experiments. The exponential functions $\varphi(x) = \exp\big\{(1-\alpha)x\big\}(\alpha \neq 1)$ meet that requirement, and accordingly, Rényi defined his entropy of order $\alpha \neq 1$ as $H_\alpha(P) = \frac{1}{1-\alpha}\log_2 \sum p_k^\alpha$.

Since $H_\alpha(P)$ converges to Shannon's entropy $H(P)$ as $\alpha \to 1$, the latter is regarded as entropy of order $\alpha = 1$. Via similar considerations, Rényi also defined I-divergence of order α (he used the term "information gain"), of which standard I-divergence is the limit as $\alpha \to 1$. He also extended his previous "dimensional entropy" results to entropy of order α.

The theory of generalized information measures initiated by Rényi's work has now an extensive literature. Since poorly motivated generalizations have also been published, it is important to note the Rényi did not endorse those. He emphasized, see his 1965 survey paper [151, article 242], that only such quantities deserve to be called information measures that can be effectively used in solving concrete problems. Rényi was able to find interesting problems whose solution involved entropy of order $\alpha \neq 1$, namely in the theory of random search (see the subsection on that topic). Later, coding problems were also found that led to Rényi entropy, see Lorain Campbell {2} and Csiszár {5}. The latter paper gives an operational characterization of Rényi's information measures (including an "order α" analogue of Shannon's mutual information, somewhat different from that suggested by Rényi) within the standard Shannon theory framework.

In 1959, Yuri Linnik presented an information-theoretic proof of the central limit theorem that intrigued Rényi. He observed that the essence of Linnik's idea was that convergence of a sequence of probability distributions P_n to a limiting distribution P may be proved by showing that the I-divergence of P_n from P converges to 0. Rényi hoped to simplify Linnik's very difficult proof by using an I-divergence of order α, say with $\alpha = 2$; this was a major cause of his interest in generalized information measures. Rényi's hope to simplify Linnik's proof did not come true, and it is still an open question under what conditions does the I-divergence of the distribution of the normalized sum of n independent random variables from the standard normal distribution converge to 0 (only the case of identically distributed summands comes close to be satisfactorily settled, see Andrew Barron's paper {1}. On the other hand, Rényi showed in his Berkeley Symposium paper [151, article 180] that the "information theoretic method" leads to a simple proof of the convergence of n-step conditional distributions of a stationary (finite state) Markov chain to the stationary distribution, and I-divergence of order α and some more general expressions are equally suitable for that purpose. The last observation motivated Csiszár {3} to introduce a general class of information-type measures of distance of probability distributions, corresponding to arbitrary convex functions f; these f-divergences turned out to have many applications in statistics, see Igor Vajda's book [188]. Rényi's followers to prove limit theorems for Markov chains via the information theoretic method include David Kendall {14} and Fritz {10}; for more recent applications of this idea, in the theory of interacting particles systems, see Liggett's book [110].

Axiomatic characterizations

Shannon's main justification for his information measures was their usefulness in communication problems, but he also showed that his entropy was uniquely characterized by certain postulates that a measure of amount of information was intuitively expected to satisfy. Later, starting with Aleksandr Jakovlevich Khinchin (Usp. Mat. Nauk, Vol. 8, 1953, pp. 3–51) and D. K. Fadeev (Usp. Mat. Nauk, Vol. 11, 1956, pp. 227–231), several mathematicians put forward axiomatic characterizations using weaker or "more natural" postulates. Rényi also contributed to this direction of research that he considered conceptually important for IT. An instructive exposition of his view appears in his survey paper [151, article 242]. He says that what he calls the axiomatic and pragmatic approaches to the problem of measuring information "are compatible and even complement each other and therefore both deserve attention. Both of the mentioned approaches may and should be used as a control of the other."

Axiomatic considerations appeared already in Rényi's first IT paper [151, article 121]. In [151, article 159], he pointed out that the key step in Fadeev's characterization (loc. cit.) was a number theoretic result that had been previously proven by Erdős {9}, and he gave a new simple proof of that result. It says that an additive number theoretic function must be equal to constant times $\log n$ if it satisfies $\lim_{n\to\infty}\big(F(n+1)-F(n)\big)=0$. Actually, the latter hypothesis may be weakened to $\liminf_{n\to\infty}\big(F(n+1)-F(n)\geq 0$, a fact later also used in a characterization of Shannon entropy, see *Zoltán Daróczy* and *Imre Kátai,* {8}.

Axiomatic characterizations play a substantial role in Rényi's Berkeley Symposium paper [151, article 180] on information measures of order α. The postulates there involve (generalized) means, and information measures are assumed to be assigned also to "incomplete distributions" (where the sum of probabilities is less than 1). When characterizing I-divergences of order α, the additivity postulate is shown to admit only means corresponding to a linear or exponential function φ, while when characterizing entropies, the same remains an unproven assumption. As reported in [151, article 242], that deficiency could be removed: Daróczy in {7} showed that entropy of order α could be satisfactorily characterized by Rényi's postulates, even without recourse to incomplete distributions.

The majority of the (substantial) contributions of Hungarian mathematicians to axiomatic characterizations of information measures (see the book of *János Aczél* and Daróczy [5]) is out of the scope of this volume. It should be noted that today this subject is not considered of primary importance for IT but, on the other hand, research in this direction has strongly contributed to the development of the theory of functional equations.

Random search

Search theory is a subject whose systematic development was initiated by Rényi. He regarded it as part of IT, which is not the prevailing view today. Still, Rényi's work on random search certainly has an IT flavor, in particular, it provides operational justification for Rényi entropies of order α.

Rényi had been fascinated by the guessing game called "twenty questions" in the US. In this game, consecutive questions (in the US, at most 20) answerable by yes or no are asked about an unknown object, in order to identify it from the answers. In Rényi's lectures, this game was a standard example to visualize the basic ideas of IT. His research about random search was motivated by his interest in what happens if the questions are selected not by a well designed strategy but at random. Rényi showed that consecutive random selections from the set of all possible questions admit identification with only slightly more questions than an optimal strategy. This visualizes a key idea of IT, the efficiency of random coding, although this fact was not emphasized by Rényi.

In 1961/62, Rényi published four papers on random search, of main interest is [151, article 193]. There, identification of elements of a set H of size n via k questions, each with r possible answers, is considered. Each question corresponds to a labeling of the elements of H by numbers randomly selected from $\{1, \ldots, r\}$ with probabilities $p_1, \ldots, p_r$, independently of each other; the answer to this question, for any fixed $x \in H$, gives the label assigned to x. The k questions identify x if for no other $y \in H$ are all k answers the same as for x. Problem: when r is fixed and $n \to \infty$, how large a k is needed in order that with a prescribed probability, either (i) a particular $x \in H$ be identified or (ii) all elements of H be identified by k random questions as above. When $P = (p_1, \ldots, p_r)$ is the uniform distribution, Rényi's result was that $k \sim \frac{\log_2 n}{\log_2 r}$ questions are needed in case (i), and $k \sim \frac{2 \log_2 n}{\log_2 r}$ in case (ii). When P is arbitrary, one would expect that $\log_2 r$ should be replaced by the entropy of P. The remarkable results was, however, that while this expectation is correct in case (i), it is Rényi entropy of order 2 rather than Shannon entropy that enters in case (ii). This was the first operational characterization of a Rényi entropy of order α, though only for $\alpha = 2$.

Later, Rényi found modified versions of the above random search problem whose solution involved entropies of other (positive integer) orders α, see his 1965 survey paper [151, article 242]. In the same year, he introduced a general model of random search in his Invited Address [151, article 249]. Here, the role of the previous random labellings of the set H is played by functions $f : H \to \{1, \ldots, r\}$ randomly selected from a given class F of such functions, thus the labels (function values $f(x)$) need not be independently chosen for each $x \in H$. Under homogeneity conditions on the function class F, general results were obtained on the probability that k questions identify a fixed element of H, or all elements of H. These, in particular cases, lead to asymptotic results similar to those mentioned above.

Information theoretic methods in statistics

A basic problem in statistics is to infer an unknown distribution P from an observed sample $X^n = (X_1, \ldots, X_n)$ where $X_1, X_2 \ldots$ are independent random variables of distribution P. If the unknown P is assumed to belong to a known family of distributions, $\{P_\theta\}$, where θ is a (scalar or vector valued) parameter ranging over a given set, one has to estimate the true value of θ from the observed sample X^n.

Information measures relevant for this problem include Fisher's information and I-divergence. Still, the natural measure of the amount of information the sample gives for the unknown parameter is Shannon's mutual information $I(X^n, \theta)$, provided one is willing to adopt the Bayesian approach of assigning a prior distribution to the parameter, necessary for the definition of $I(X^n, \theta)$. The first to study the asymptotic behavior of $I(X^n, \theta)$ as $n \to \infty$ was Rényi, in the special case when θ had a finite number of possible values. Starting in 1964, Rényi treated this issue in 9 papers; the strongest results appear in [151, article 288] and [151, article 328]. He related this problem to that of the asymptotic behavior of the error probability of the Bayesian estimator of θ (called by him the "standard decision" Δ_n) and studied both problems in parallel.

Rényi proved that $I(X^n, \theta)$ converges to the entropy of θ or – equivalently – the missing information (Rényi's term for the conditional entropy of θ given X^n) converges to 0 exponentially fast, and so does the error probability $P(\Delta_n \neq \theta)$, with the same exponent. When θ had two possible values only, Rényi gave an exact asymptotic formula for $P(\Delta_n \neq \theta)$, using large deviations techniques. He also considered the case of (independent but) not identically distributed observations, and gave an upper bound to the missing information in terms of Hellinger integrals. This result is related to Kakutani's theorem that for two sequences $\{\mu_n\}$ and $\{\nu_n\}$ of probability measures, their infinite products are mutually absolutely continuous or singular according as the infinite product of the Hellinger integrals $\lambda_n = \int \sqrt{d\mu_n \, d\nu_n}$ is positive or 0.

Rényi's work gave substantial impetus to studying the interplay of statistics and IT. Bounds to error probability in terms of Shannon's and more general information measures are too numerous to cite here. Rényi's asymptotic formula for $P(\Delta_n \neq \theta)$ when θ has two possible values easily extends to θ with any finite number of values, see independent work of Vajda (Proc. Coll. Inform. Theory, Debrecen 1967, J. Bolyai Math. Society, Budapest, 1969, pp. 489–501). Rényi's result related to Kakutani's theorem actually gives an information theoretic proof of one half of that theorem, namely that $\Pi_{n=1}^{\infty} \lambda_n = 0$ implies singularity. An information theoretic proof of the other half was given by *Tibor Nemetz* {15}. Extensions of the study of the asymptotic behavior of $I(X^n, \theta)$ to the case of a continuous parameter θ, turned out of substantial interest for the theory of universal coding and Bayesian statistics, see Bertrand Clarke and Barron, J. Statist. Plan. Inference, Vol. 41, 1999, pp. 37–60. If θ is a k-dimensional parameter, $I(X^n, \theta)$ is asymptotically equal to $\frac{k}{2}\log_2 n$ plus a constant that depends on the prior distribution of θ (subject to regularity conditions); the constant is smallest for the so-called Jeffreys prior which is thereby distinguished as "least informative."

Vincze also devoted several papers to the interplay of IT and statistics, in particular to information-type measures of distance of probability distributions that belong to the class of f-divergences, see the subsection about Rényi informations. Members of this class of statistical significance include Hellinger distance and χ^2-distance, corresponding to $f(t) = 1 - \sqrt{t}$ and $f(t) = (t-1)^2$. Puri and Vincze in {17} introduced distances denoted by $I_N(P, Q)$ that correspond to $f(t) = \frac{1}{2}\frac{|t-1|^N}{(t+1)^{N-1}}$, $N \geq 1$. Their main result was that two sequences of probability distributions $\{P_n\}$ and $\{Q_n\}$ are mutually contiguous if and only if $I_{N_n}(P_n, Q_n) \to 0$ for all sequences of numbers $N_n \to \infty$.

Kafka, Ferdinand Österreicher and Vincze in {13} studied the problem for what f-divergences was some power of them a metric, and then what was the smallest such power. Previously, Csiszár and *János Fischer* in {6} showed that for $f(t) = 1 - t^\alpha$, $0 < \alpha < 1$, the corresponding f-divergence raised to the power $\min(\alpha, 1-\alpha)$ satisfied the triangle inequality (but was not a metric, for lack of symmetry, except if $\alpha = 1/2$, the case of Hellinger distance). The results of Kaffka, Österreicher and Vincze include that $\big(I_N(P, Q)\big)^{\frac{1}{N}}$ is a metric, the f-divergence corresponding to $f(t) = 1 + t^\alpha - t^{1-\alpha}$ raised to the power $\min(\alpha, 1-\alpha)$ is a metric, and so is the square-root of an f-divergence that gives the perimeter of the risk set for testing P against Q, studied earlier by Österreicher {16}; in neither case is any smaller power appropriate.

References

[5] Aczél, János – Daróczy, Zoltán, *On Measures of Information and their Characterizations* Academic Press (New York, 1975).

[98] Kullback, Solomon, *Information Theory and Statistics,* Wiley (New York, 1959), Dover (New York, 1978).

[110] Liggett, Thomas M., *Interacting Particle Systems,* Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen, Band 276, Springer-Verlag (New York, 1985).

[151] Rényi, Alfréd, *Selected Papers,* ed. Pál Turán, Akadémiai Kiadó (Budapest, 1976).

[152] Rényi, Alfréd, *Probability Theory,* Translated by László Vekerdi, North-Holland Series in Applied Mathematics and Mechanics, Vol. 10, North-Holland Publishing Company (Amsterdam–London); American Elsevier Publishing Co., Inc. (New York, 1970).

[154] Reza, Fazlollah M., *An Introduction to Information Theory,* McGraw-Hill (New York, 1961).

[188] Vajda, Igor, *Theory of Statistical Inference and Information,* Kluwer Academic (Boston, 1989).

[192] Wald, Ábrahám, *Sequential Analysis,* John Wiley and Sons (New York) – Chapman and Hall (London, 1947).

{1} A. Barron, Antropy and the central limit theorem, *Annals of Probability,* **14** (1986), 336–342.

{2} L. Campbell, A coding theorem and Rényi's entropy, *Information and Control,* **8** (1965), 423–429.

{3} I. Csiszár, Eine inrofmationtheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen ketten, *Publ. Math. Inst. Hungar. Acad. Sci.,* **8** (1963), 85–108.

{4} I. Csiszár, Generalized entropy and quantization problems, *Trans. Sixth Prague Conference on Inform. Theory, etc.,* 1971, Academia (Praha, 1973), 299–318.

{5} I. Csiszár, Generalized cutoff rates and Rényi's information measures, *IEEE Trans. Inform. Theory,* **41** (1995), 26–34.

{6} I. Csiszár and J. Fischer, Informationsentfernungen im Raum der Wahrscheinlichkeitsverteilungen, *Publ. Math. Inst. Hungar. Acad. Sci.,* **7** (1962), 159–182.

{7} Z. Daróczy, Über Mittelwerte und Entropien vollstandiger Wahrscheinlichkeitsverteilungen, *Acta Math. Sci. Hungar.,* **15** (1964), 203–210.

{8} Z. Daróczy and I. Kátai, Additive zahlentheoretische Funktionen und das Mass der information, *Ann. Univ. Sci. Budapest, Sec. Math.,* **13** (1970), 83–88.

{9} P. Erdős, On the distribution function of additive functions, *Annals of Math.,* **17** (1946), 1–20.

{10} J. Fritz, An information-theoretical proof of limit theorems for reversible Markov processes, *Trans. Sixth Prague Conference on Inform. Theory, etc.,* 1971. Academia (Praha, 1973), 183–197.

{11} J. Fritz, An approach to the entropy of point processes, *Periodica Math. Hungar.,* **3** (1973), 73–83.

{12} P. Gács, Hausdorff-dimension and probability distribution, *Periodica Math. Hungar.,* **3** (1973), 59–71.

{13} P. Kafka, F. Österreicher and I. Vincze, On powers of f-divergences defining a distance, *Studia Sci. Math. Hungar.,* **26** (1991), 415–422.

{14} D. Kendall, Information theory and the limit-theorem for Markov chains and processes with a countable infinity of states, *Annals Inst. Statist. Math.,* **15** (1963), 137–143.

{15} T. Nemetz, Equivalence-orthogonality dichotomies of probability measures, *Limit Theorems of Probability Theory,* Colloquia Math. Soc. J. Bolyai, Vol. 11, North Holland (1975), 183–191.

{16} F. Österreicher, The construction of least favourable distributions is traceable to a minimal perimeter problem, *Studia Sci. Math. Hungar.,* **17** (1982), 341–351.

{17} M. Puri and I. Vincze, Measure of information and contiguity, *Statistics and Probability Letters,* **9** (1990), 223–228.

{18} M. Rudemo, Dimension and entropy for a class of stochastic processes, *Publ. Math. Inst. Hungar. Acad. Sci.,* **9** (1964), 73–87.

{19} C. Shannon, A mathematical theory of communication, *Bell System Technical Journal,* **27** (1948), 379–423 and 623–656.

{20} C. Shannon, Communication in the presence of noice, *Proc. IRE,* **37** (1949), 10–21.

Imre Csiszár

Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
P.O.B. 127
1364 Budapest
Hungary

`csiszar@renyi.hu`