István Vincze (1912–1999) and his contribution to lattice path combinatorics and statistics

Endre Csáki

Alfred Rényi Institute of Mathematics, Hungarian Academy of Sciences, P.O. Box 127, H-1364 Budapest, Hungary

Received 30 October 2002; received in revised form 31 March 2003; accepted 7 February 2005
Available online 22 March 2005

Abstract

A brief account of the life and work of István Vincze, a prominent Hungarian statistician, is given. His contributions in various topics are discussed. They include empirical distribution, Kolmogorov–Smirnov statistics, information theory, Cramér–Fréchet–Rao inequality, estimation of density, and a characterization problem.

MSC: Primary 60C05; Secondary 30D20; 60G50; 62B10; 62F10; 62G05; 62G30; 94A17

Keywords: Empirical distribution; Power function; Entropy; Cramér–Fréchet–Rao inequality; Density estimation; Entire function

1. Introduction

István Vincze was born in Szeged, Hungary, on February 26, 1912. After his graduation from the University of Szeged in 1935, he worked for a Hungarian Insurance Company until 1945. The second world war interrupted his career. After the war, he worked for the Ministry of Education until 1950. Then, he was invited by the late Alfréd Rényi to join an Institute, whose main duty was set to do both theoretical and applied Mathematics. In
this way, he became one of the founders of the Mathematical Institute of the Hungarian Academy, whose director was Alfréd Rényi. Vincze was the Head of the Statistics Department until his retirement in 1980, and during 1950–1964 he also served as Deputy Director of the Institute. He was also a Professor in Statistics at the Eötvös Loránd University, Budapest. I had the privilege to be one of his numerous students in Statistics. He was considered as one of the main experts in both Theoretical and Applied Statistics in Hungary and also all over the world. Although in the early stage of his research activity he was interested in Geometry, on which he wrote several papers, including joint papers with Erdős, he has made significant contributions to several branches of Statistics, such as Quality Control, Nonparametric Statistics, Empirical distributions, Cramér–Rao inequality, Information Theory, etc. He is the author of more than 100 research papers and 10 books.

He was awarded a number of honors in his life, including the Hungarian State Prize in 1966 and Gauss Ehrenplakette in 1978.

Except for the last two years of his life, he was very active even when he was over 80. He worked regularly in the Mathematical Institute, gave seminar talks, participated in conferences, such as Probastat, Bratislava, 1991 and 1994, Stochastic Modeling and Lattice Path Combinatorics, Delhi, 1994, Stability Problems, Kazan, 1995, Approximation Theory, Budapest, 1995, Statistical Conference, Poland, 1996. He was invited to the Combinatorial Methods Conference, Hamilton, Canada, 1997. He wrote a paper for the occasion (see Vincze and Tóróš, 1997), but an unfortunate accident prevented him from participating.

Professor István Vincze visited many universities and institutes all over the world. He spent several months in China, GDR, USA, Canada, Argentina, etc. He was invited as speaker to several conferences, including three Berkeley Symposiums: 1960, 1965, 1970. He also organized a number of conferences: European Meeting of Statisticians in Budapest, 1972, Nonparametric Statistical Inference in Budapest, 1980, Pannonian Symposiums on Mathematical Statistics in Bad Tatzmannsdorf in 1979, 1981 and 1983, and in Visegrád, Hungary, 1982. He was the director of the Unesco courses on Probability and Statistics, held in the Mathematical Institute, Budapest in 1964 and 1968.

Professor Vincze was a very kind man, and his hospitality was legendary. He would walk with his guests through Budapest an entire day to show them the most important tourist attractions and serve as a real guide to explain the history of Hungary attached to a particular building and place. He was physically vigorous all his life.

István Vincze will be remembered by the statistical community for his warmth, humanity and friendliness.

In this paper, we summarize the most important contributions of Professor István Vincze in the following areas, focusing mainly on the first topic, but mentioning briefly his contributions in other subjects as well:

- empirical distribution, random walk, lattice paths,
- information theory,
- Cramér–Fréchet–Rao inequality,
- estimation of density and its derivatives,
- a characterization problem.
2. Empirical distribution, random walk, lattice paths

Professor Vincze was the main contributor to the theory of empirical distributions and random walks (lattice paths), which were among his favorite topics. Consider a random sample

\[(X_1, X_2, \ldots, X_n)\]

of size \(n\), coming from a population with (theoretical) distribution function \(F(x) = P(X_1 \leq x)\). The empirical or sample distribution function is defined by

\[F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq x\},\]

where \(I(A)\) stands for the indicator of the event \(A\). Empirical distribution functions are widely used in statistics, nonparametric statistics in particular. In the two-sample case, Gnedenko and Korolyuk (1951) developed a method based on random walk models. Let \((X_1, X_2, \ldots, X_n)\) and \((Y_1, Y_2, \ldots, Y_m)\) be two samples coming from continuous distributions. Let \(F(x)\) and \(G(x)\) resp. be their theoretical distribution functions and let \(F_n(x)\) and \(G_m(x)\) resp. be their empirical distribution functions. Testing the null hypothesis \(H_0 : F(x) = G(x)\), a number of statistics have been investigated and their distributions, limiting distributions and other characteristics have been determined in the statistical literature. The idea of Gnedenko and Korolyuk was as follows: let

\[Z_1^* < Z_2^* < \cdots < Z_{n+m}^*\]

denote the order statistics of the union of the two samples and define

\[\theta_i = \begin{cases} +1 & \text{if } Z_i^* = X_j \text{ for some } j, \\ -1 & \text{if } Z_i^* = Y_j \text{ for some } j, \end{cases} i = 1, 2, \ldots, n + m.\]

Put

\[S_0 = 0, \quad S_i = \theta_1 + \cdots + \theta_i, \quad i = 1, 2, \ldots, n + m.\]

Then \((S_0, S_1, \ldots, S_{n+m})\) is a random walk path with \(S_{n+m} = n - m\) and under \(H_0\) each of them has the same probability. This idea of Gnedenko and Korolyuk enables one to determine the distributions of certain statistics by reducing the problems to combinatorial enumeration.

In a series of papers, Professor Vincze and his collaborators presented a number of results on this subject. His first result concerns the joint distribution of the maximum and its location in the case of \(n = m\). Define

\[B_n^+ = n \max_{(x)} (F_n(x) - G_n(x)) = \max_{1 \leq i \leq 2n} S_i\]

and let \(R_n^+\) be the first index \(i\) for which this maximum is achieved.
Moreover, put

\[ B_n = n \max_{(x)} |F_n(x) - G_n(x)| = \max_{1 \leq i \leq 2n} |S_i| \]

and let \( R_n \) be the first index \( i \) for which this maximum is achieved.

Then, under \( H_0 \), Vincze (1958) showed

\[ P(B_n^+ = k, R_n^+ = r) = \frac{k(k + 1)}{r(2n - r + 1)} \frac{r^{(r+k)/2}}{(2n-r+k)/2} \frac{1}{n-r+k/2} \left( \frac{2n}{n} \right), \]

\( k = 1, 2, \ldots, n; \ r = k, k + 2, \ldots, 2n - k. \)

Concerning the joint limiting distribution, it was shown that

\[ \lim_{n \to \infty} P \left( \frac{B_n^+}{\sqrt{2n}} < y, \frac{R_n^+}{n} < z \right) = \sqrt{\frac{2}{\pi}} \int_0^y \int_0^z u^2 (v(1-v))^{3/2} \exp \left( - \frac{u^2}{v(1-v)} \right) \, du \, dv. \]

Furthermore,

\[ P(B_n = k, R_n = r) = \frac{2A_r^{(k)} A_{2n-r+1}^{(k+1)}}{(2n)} \]

with

\[ A_r^{(k)} = \sum_{j=0}^{\infty} (-1)^j (2j + 1)k \frac{r}{r+k/2+jk} \]

and

\[ \lim_{n \to \infty} P \left( \frac{B_n}{\sqrt{2n}} < y, \frac{R_n}{n} < z \right) = \sqrt{\frac{8}{\pi}} \int_0^y \int_0^z f(u, v) f(u, 1-v) \, du \, dv, \]

where

\( f(y, z) = \frac{y}{z^{3/2}} \sum_{j=0}^{\infty} (-1)^j (2j + 1) e^{-((2j+1)^2 y^2)/2z}. \)

Reimann and Vincze (1960) studied the case of different sample sizes. Define

\[ B_{n,m}^+ = \max_{(x)} (n F_n(x) - m G_m(x)) = \max_{1 \leq i \leq n+m} S_i \]

and let \( R_{n,m}^+ \) be the first index \( i \) for which this maximum is achieved.
Furthermore, put

\[
B_{n,m} = \max_{(x)} \left| nF_n(x) - mG_m(x) + \frac{m-n}{2} \right| - \frac{m-n}{2}
\]

\[
= \max_{1 \leq i \leq n+m} \left| S_i + \frac{m-n}{2} \right| - \frac{m-n}{2},
\]

and let \( R_{n,m} \) be the first index \( i \) for which this maximum is achieved. Let \( m > n \). It was shown that

\[
P(B_{n,m}^+ = k) = \frac{2k + 1 + m - n}{m + k + 1} \left( \frac{m+n}{n-k} \right)
\]

and

\[
P(B_{n,m} = k) = \frac{1}{\binom{m+n}{n}} \sum_{j=-\infty}^{\infty} \left( \binom{m+n}{m+j} - \binom{m+n}{m+k+j} \right)
\]

\[
= \frac{2^{m+n+1}}{s \binom{m+n}{n}} \sum_{l=1}^{\infty} \cos^{m+n} \frac{l\pi}{s} \sin \frac{k\pi}{s} \sin \frac{(s-k)l\pi}{s},
\]

with \( s = 2k + m - n \). Joint distributions of \((B, R)\) and limiting distributions were also given.

These Reimann–Vincze statistics are different from the usual Kolmogorov–Smirnov statistics \( \sup_x (F_n(x) - G_m(x)) \) or \( \sup_x |F_n(x) - G_m(x)| \), but have the advantage of easier computations of their distributions. Koul and Quine (1974) have shown that when the sample sizes are slightly different only, then the Bahadur efficiency of Reimann–Vincze statistics relative to the Kolmogorov–Smirnov statistics is 1.

Vincze (1959, 1963), proposed the use of generating functions to determine the above distributions and joint distributions. It was shown that

\[
\sum_{n=k}^{\infty} \sum_{r=k}^{2n-k} \binom{2n}{n} P(B_n^+ = k, R_n^+ = r) v^r w^n
\]

\[
= \frac{2^{2k+1} v^k w^k}{(1 + \sqrt{1 - 4v^2w})^k (1 + \sqrt{1 - 4w})^{k+1}}
\]

and

\[
\sum_{n=k}^{\infty} \sum_{r=k}^{2n-k} \binom{2n}{n} P(B_n = k, R_n = r) z^{r-k} w^{n-k}
\]

\[
= 2 \frac{(1 + \omega(w))^{k+1}(1 + \omega(z^2 w))^k}{(1 + (\omega(w))^{k+1})(1 + (\omega(z^2 w))^k)},
\]

with

\[
\omega(z) = \frac{1 - \sqrt{1 - 4z}}{1 + \sqrt{1 - 4z}} = \frac{4z}{(1 + \sqrt{1 - 4z})^2}.
\]
Now let 
\[(X^*_1 < X^*_2 < \cdots < X^*_n), \quad (Y^*_1 < Y^*_2 < \cdots < Y^*_n)\]
denote the ordered samples. Then
\[
\gamma_n = \sum_{i=1}^{n} I\{X^*_i > Y^*_i\} = \frac{1}{2} \sum_{i=1}^{2n} (I[S_i > 0] + I[S_{i-1} = +1, \ S_i = 0])
\]
is the so-called Galton statistics. Chung and Feller (1949) showed that \(\gamma_n\) is uniformly distributed, i.e.,
\[
P(\gamma_n = g) = \frac{1}{n + 1}, \quad g = 0, 1, 2, \ldots, n.
\]
Csáki and Vincze (1961) considered the number of crosses
\[
\lambda_n = \sum_{i=1}^{2n-1} I\{S_i = 0, \ S_{i-1}S_{i+1} < 0\}
\]
and showed that
\[
P(\lambda_n = \ell - 1) = \frac{2\ell}{n} \binom{2n}{\ell - 1}, \quad \ell = 1, 2, \ldots, n
\]
and
\[
\lim_{n \to \infty} P(\lambda_n < y\sqrt{2n}) = 1 - e^{-2y^2}, \quad y \geq 0.
\]
The joint exact and limiting distributions of \((\gamma_n, \lambda_n)\) were also given, namely,
\[
P(\gamma_n = g, \lambda_n = \ell - 1) = \frac{1}{(2n/n)^2} \frac{\ell^2}{2g(n-g)} \left( \frac{2g}{g - \ell/2} \right) \left( \frac{2n-2g}{n-g-\ell/2} \right)
\]
for \(\ell\) even, and
\[
P(\gamma_n = g, \lambda_n = \ell - 1) = \frac{1}{(2n/n)^2} \frac{\ell^2 - 1}{4g(n-g)} \left( \left( \frac{2g}{g - (\ell + 1)/2} \right) \left( \frac{2n-2g}{n-g-(\ell-1)/2} \right) + \left( \frac{2g}{g - (\ell - 1)/2} \right) \left( \frac{2n-2g}{n-g-(\ell+1)/2} \right) \right)
\]
for \(\ell\) odd. For the joint limiting distribution, it was shown that
\[
\lim_{n \to \infty} P(\gamma_n \leq zn, \lambda_n \leq y\sqrt{2n}) = \sqrt{\frac{2}{\pi}} \int_0^y \int_0^z \frac{u^2}{(v(1-v))^{3/2}} \exp\left(-\frac{u^2}{2v(1-v)}\right) \, du \, dv.
\]
Another use of the generating function method is found in Csáki and Vincze (1963b), where the joint distribution of the maximum and the number of crosses was given in the form
\[
\sum_{n=1}^{\infty} \binom{2n}{n} \mathbb{P} \left( \max_{x} |F_n(x) - G_n(x)| = \frac{k}{n}, \; \lambda_n = \ell - 1 \right) z^n \\
= 2 \left( \frac{w - w^k}{1 - w^{k+1}} \right)^{\ell}, \quad \ell, k = 1, 2, \ldots,
\]
where
\[
w = \frac{1 - \sqrt{1 - 4z}}{1 + \sqrt{1 - 4z}}, \quad |z| < \frac{1}{4}.
\]

These and related results were presented by Vincze on a number of occasions in conferences, including the 4th, 5th and 6th Berkeley Symposium on Mathematical Statistics and Probability.

Vincze (1961) considered two-dimensional samples \((X_i^{(1)}, X_i^{(2)})\) and \((Y_i^{(1)}, Y_i^{(2)})\), \(i = 1, 2, \ldots, n\) from distributions having theoretical distribution functions \(F(x, y)\) and \(G(x, y)\), resp. Empirical distribution functions are denoted by \(F_n(x, y)\) and \(G_n(x, y)\), resp. Testing \(F = G\), he proposed to choose \(\eta = y\) randomly according to the distribution function \(H(y) = F(\infty, y)\) and consider the maximum deviation between \(F_n(x, \eta)\) and \(G_n(x, \eta)\). The following distributions were determined:
\[
P \left( \max_{(x)} (F_n(x, \eta) - G_n(x, \eta)) < \frac{k}{n} \right) \\
= \frac{1}{(2n + 1)} \binom{2n}{n} \sum_{i=0}^{n} \min(n, i+k) \sum_{j=\max(0, i-k)}^{n} \binom{2n-i-j}{n-i} \left( \binom{i+j}{i} - \binom{i+j}{i-k} \right)
\]

and
\[
P \left( \max_{(x)} |F_n(x, \eta) - G_n(x, \eta)| < \frac{k}{n} \right) \\
= \frac{1}{(2n + 1)} \binom{2n}{n} \sum_{i=0}^{n} \min(n, i+k) \sum_{j=\max(0, i-k)}^{n} \binom{2n-i-j}{n-i} \sum_{h=-\infty}^{\infty} (-1)^k \binom{i+j}{i+hk}.
\]

Vincze’s idea in determining joint distributions was to construct tests based on a pair of statistics (instead of one single statistic) in order to improve the power of the tests. Vincze (1965, 1967, 1968a) studied the power function of several tests based on the empirical distribution function. He computed the power function of the two-sample Smirnov test and showed by numerical examples that the power can be increased by using a pair of statistics instead of one statistic. In particular, he considered the maximum and its location as a pair of statistics.

In Vincze (1970, 1972), several problems are treated concerning Kolmogorov–Smirnov statistics. Among others an explicit formula is given for the two-sample case for discontinuous random variables and a discussion is given in the two-dimensional case.
Finally, we mention several results in Csáki and Vincze (1961, 1963a, 1964), concerning equivalence relations, proved by bijections. Define the following quantities:

\[ \lambda_n' = \sum_{i=1}^{2n-1} I\{S_{i-1} = 0, S_i = +1\}, \]
\[ \pi_n = \sum_{i=1}^{2n} I\{S_i > 0\}, \]
\[ \tau_\ell = \min\{i : S_i = \ell\}. \]

Then we have

\[ \{\lambda_n = \ell - 1, S_1 = +1\} \iff \{\tau_2\ell = 2n\}, \]
\[ \{B_n^+ = \ell, R_n^+ = r\} \iff \{\lambda_n' = \ell, \pi_n = r\}. \]

Here \( \{\cdots\} \iff \{\cdots\} \) means that there is a bijection between the two sets of random walk paths.

These relations are valid in the case when \( S_{2n} = 0 \). Now consider the general case, i.e., no restriction on the terminal point. Define

\[ 2\gamma_{2n}'^{(2k)} = \sum_{i=1}^{2n} (I\{S_i > 2k\} + I\{S_{i-1} = 2k + 1, S_i = 2k\}). \]

Then

\[ \{\gamma_{2n}'^{(2k)} = g\} \iff \{S_{2n-2g} = S_{2n} = 2k\}. \]

The last relation implies the following distribution result:

\[ \mathbb{P}(\gamma_{2n}'^{(2k)} = g) = \frac{1}{2^{2n}} \binom{2g}{g} \binom{2n - 2g}{n - g + k}, \quad g = 1, \ldots, n - k. \]

In case \( k = 0 \), we recover the finite arcsine law of Chung and Feller (1949).

### 3. Information theory

Vincze (1960) gave an interpretation of the \( I \)-divergence as below, concerning the information of a continuous random variable relative to “the distribution of our interest”.

The entropy of a system of events \( A_1, A_2, \ldots, A_n \), in the case \( P(A_i) = p_i \) \((i = 1, 2, \ldots, n)\), is \( E_n = \sum_{i=1}^{n} p_i \log(1/p_i) \). Consider the quantity

\[ I_n = \log n - E_n = \sum_{i=1}^{n} p_i \log np_i, \quad 0 \leq I_n \leq \log n, \]

which is called the “information of the system of events”. Let \( \xi \) be a continuous random variable with distribution function \( F(x) \) and density function \( F'(x) = f(x) > 0 \) for any value
of $x$. Vincze introduces a distribution function $\phi$ called the “distribution of our interest”. Consider a system of divisions of the real line such that the points of division are regarded as the quantiles of a distribution function, that is, $\phi(x^{(n)}_k) = k/n$, $k = 1, 2, \ldots, n - 1$; $n = 1, 2, \ldots$, $\phi(-\infty) = 0$, $\phi(\infty) = 1$. For each $n$, the information of the system of discrete events $A_i^{(n)} = \{x_i^{(n)} \leq \xi < x_i^{(n)}\}$ ($i = 1, 2, \ldots, n$) is

$$I_{n,\phi}(\xi) = \sum_{i=1}^{n} (F(x_i^{(n)}) - F(x_{i-1}^{(n)})) \log(n(F(x_i^{(n)}) - F(x_{i-1}^{(n)}))),$$

or, after the substitution of $1/n = \phi(x_i^{(n)}) - \phi(x_{i-1}^{(n)})$,

$$I_{n,\phi}(\xi) = \sum_{i=1}^{n} (F(x_i^{(n)}) - F(x_{i-1}^{(n)})) \log \frac{F(x_i^{(n)}) - F(x_{i-1}^{(n)})}{\phi(x_i^{(n)}) - \phi(x_{i-1}^{(n)})}.$$

Assuming that $\phi'(x) = \varphi(x) > 0$ for all real $x$, by a limiting process one obtains the relation

$$\lim_{n \to \infty} I_{n,\phi}(\xi) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{\varphi(x)} \, dx = I_{\phi}(\xi),$$

which is the so-called $I$-divergence (Kullback, 1959). The expression $I_{\phi}(\xi)$ can be regarded as the information of the random variable relative to the distribution $\phi$ of our interest.

For another use of $I$-divergence as treated in Vincze (1974), we state the following result of Sanov (1957): let $X$ be a random variable whose distribution function is $\phi(x)$ and $F(x)$ be any distribution function such that the Borel measures $\mu_\phi$ and $\mu_F$ induced, respectively, by $\phi(x)$ and $F(x)$ are equivalent. Let $\phi_N(x)$ be the empirical distribution function of $X$ after a large number $N$ of independent trials; then for every $\varepsilon > 0$

$$P \left( \sup_x |\phi_N(x) - F(x)| < \varepsilon \right) \approx \exp(-NI),$$

where $I = \int \log[dF(x)/d\phi(x)] \, dF(x)$ is the $I$-divergence. On the basis of this result, Vincze (1974) gives a correct formulation of the maximum-probability principle valid for both continuous and discrete random quantities. The principle consists in finding the distribution function $F(x)$ which minimizes the $I$-divergence under the constraints $\int x \, dF(x) = m \neq \int x \, d\phi(x)$ and $\int dF(x) = 1$. It is then shown that the $I$-divergence is the natural extension to continuous random variables of Shannon’s formula for discrete entropy. Consequently, the maximum-probability principle provides an information-theoretical foundation of statistical mechanics, as suggested by Jaynes (1957), who considered only the discrete case.

Vincze (1975) also extends the maximum-probability principle to the case where the components of the system are not statistically independent, and for illustration he discusses the derivation of the Bose–Einstein and the Fermi–Dirac distributions.


Let \( X = (X_1, X_2, \ldots, X_n) \) be a sample from a distribution having (joint) density \( p(x; \theta) = p(x_1, x_2, \ldots, x_n; \theta) \) with respect to a measure \( \mu \), where \( \theta \) is a parameter. Let \( t(X) \) be an unbiased estimator of \( g(\theta) \), i.e. \( E_\theta(t(X)) = g(\theta) \). Cramér (1946), Fréchet (1943) and Rao (1945) concluded the following inequality:

\[
\text{Var}_\theta(t(X)) \geq \frac{(g'(\theta))^2}{I(\theta)},
\]

with

\[
I(\theta) = \int \left( \frac{\partial p}{\partial \theta} \right)^2 p(x; \theta) \, dx.
\]

For fixed \( \theta, \theta' \), Vincze (1979, 1981) considered the mixture

\[
p_\alpha = p_\alpha(x; \theta, \theta') = (1 - \alpha) p(x; \theta) + \alpha p(x; \theta'), \quad 0 < \alpha < 1
\]

with \( \alpha \) being a new parameter. Then

\[
\hat{\alpha} = \frac{t(X) - g(\theta)}{g(\theta') - g(\theta)}
\]

is an unbiased estimator of \( \alpha \).

It follows that

\[
\text{Var}_\alpha(\hat{\alpha}) \geq \frac{1}{J_\alpha(\theta, \theta')},
\]

where

\[
J_\alpha(\theta, \theta') = \int \frac{(p(x; \theta') - p(x; \theta))^2}{p_\alpha(x; \theta, \theta')} \, d\mu.
\]

Then

\[
(1 - \alpha) \text{Var}_\theta(t(X)) + \alpha \text{Var}_{\theta'}(t(X)) \geq \frac{1}{J_\alpha(\theta, \theta')} - \alpha(1 - \alpha)
\]

and in the case when \( \text{Var}_\theta(t(x)) \) does not depend on \( \theta \), Vincze concluded the following lower bound:

\[
\text{Var}(t(X)) \geq \sup_{\alpha} \sup_{\theta'} \alpha(1 - \alpha)(g(\theta') - g(\theta))^2 \left( \frac{1}{\alpha(1 - \alpha) J_\alpha} - 1 \right).
\]

In certain cases this gives a reasonably good bound. This problem was further investigated by Puri and Vincze (1985), Govindarajulu and Vincze (1989) and Vincze (1992, 1996). It was shown among others that for the translation parameter of the uniform distribution this lower bound is of order \( n^{-2} \), which is attainable.
5. Estimation of density and its derivatives

Let \( f(x) \) be a density on the interval \( (a, b) \) and for positive integers consider a partition \( a = x_0^{(n)} < x_1^{(n)} < x_2^{(n)} < \cdots < x_n^{(n)} = b \). Put

\[
s_k^{(n)} = \frac{\int_{x_k^{(n)}}^{x_{k+1}^{(n)}} tf(t) \, dt}{\int_{x_k^{(n)}}^{x_{k+1}^{(n)}} f(t) \, dt}, \quad k = 0, 1, 2, \ldots, n - 1.
\]

Rényi (1952, personal communication) raised the question of whether \( f(x) \) can be determined if for each \( n \) we know \( \{s_0^{(n)}, s_1^{(n)}, s_2^{(n)}, \ldots, s_{n-1}^{(n)}\} \) for a partition, such that

\[
\lim_{n \to \infty} \max_{1 \leq i \leq n} (x_i^{(n)} - x_{i-1}^{(n)}) = 0.
\]

Vincze (1954) answered this question in the affirmative. His idea was to show that

\[
\frac{s(u, v) - (u + v)/2}{(v - u)^2} \to \frac{1}{12} \frac{f'(x)}{f(x)}, \quad \text{when } u, v \to x,
\]

where

\[
s(u, v) = \frac{\int_u^v tf(t) \, dt}{\int_u^v f(t) \, dt}.
\]

This was extended by Gupta and Vincze (1991) as follows:

\[
2^r(2r + 1)!! \frac{\int_u^v L_r(t; u, v) f(t) \, dt}{(v - u)^{r+1}} \to f^{(r)}(x), \quad \text{when } u, v \to x,
\]

where \( f^{(r)} \) denotes the \( r \)th derivative of \( f \) and \( L_r(t; u, v) \) denotes the Legendre polynomial of degree \( r \) belonging to the interval \( (u, v) \) normalized such that

\[
\int_u^v L_r^2(t; u, v) \, dt = \frac{v - u}{2r + 1}.
\]

Now suppose we want to estimate

\[
I = \int_a^b \psi \left( \frac{f'(x)}{f(x)} \right) f(x) \, dx,
\]

where \( \psi(y) \) is a given function. For example, \( \psi(y) = y^2 \).

Assume that \( (X_1, \ldots, X_N) \) is a random sample taken from a population with the distribution function having the density \( f(x) \). For a partition \( x_0 = a < x_1 < x_2 < \cdots < x_n = b \),
introduce the following notations:

\[ v_i = \sum_{j=1}^{N} I\{x_{i-1} \leq X_j < x_i\}, \]

\[ \bar{X}_{(i)} = \frac{1}{v_i} \sum_{j=1}^{N} X_j I\{x_{i-1} \leq X_j < x_i\}, \]

\[ m_i = \frac{x_i - x_{i-1} + x_i}{2}, \quad d_i = \frac{1}{\sqrt{12}}(x_i - x_{i-1}). \]

Csáki and Vincze (1977) showed that under certain regularity conditions,

\[ I_n = \sum_{i=1}^{n} \psi \left( \frac{\bar{X}_{(i)} - m_i}{d_i^2} \right) \]

is a consistent estimator of \( I \), as \( N \to \infty \).

A related question was treated in Csáki and Vincze (1978) as follows. Under the previous notation for \( \bar{X}_{(i)} \), consider

\[ \bar{z}_n^2 = \sum_{i=1}^{n} \left( \frac{\bar{X}_{(i)} - E_i}{\sigma_i} \right)^2 v_i, \]

where

\[ E_i = \mathbb{E}(X_1 \mid x_{i-1} \leq X_1 < x_i), \]
\[ \sigma_i^2 = \text{Var}(X_1 \mid x_{i-1} \leq X_1 < x_i). \]

It was shown that (for fixed \( n \)), as \( N \to \infty \), the limiting distribution of the above defined \( \bar{z}_n^2 \) statistics is \textit{chi-square} with \( n \) degrees of freedom. This provides an alternative method for a goodness of fit test instead of the usual Pearson’s chi-square test.

6. A characterization problem

Rényi and Vincze posed the following question: let

\[ f(t) = 1 + a_1 t + a_2 t^2 + \cdots + a_n t^n + \cdots \]

be an entire function. Suppose that, on the one hand,

\[ p_i = \frac{a_i t^i}{f(t)}, \quad i = 1, 2, \ldots \quad (6.1) \]

is a probability distribution for all fixed \( t > 0 \), i.e.,

\[ \sum_{i=0}^{\infty} \frac{a_i t^i}{f(t)} = 1, \quad t > 0, \]
and, on the other hand, for each $i = 0, 1, 2, \ldots$

$$\frac{a_i t^i}{f(t)}$$

is a density, i.e.,

$$\int_0^\infty \frac{a_i t^i}{f(t)} \, dt = 1, \quad i = 0, 1, 2, \ldots$$

Is it true that $f(t) = e^t$?

It is clear that for $f(t) = e^t$, $a_i = 1/i!$, (6.1) is the Poisson distribution with parameter $t$, while (6.2) is a gamma density. The converse, i.e. to show that $f(t) = e^t$ is the only solution, proved to be a rather hard problem. This (open) problem was also mentioned in a book by Hayman (1967).

In attacking this problem, Vincze and his coauthors (Hayman and Vincze, 1978, 1979; Hall and Vincze, 1981; Vincze, 1984, 1988; Csordás and Vincze, 1992), though they did not solve the problem completely, made significant steps toward the solution and obtained many interesting side results. It was shown in Hayman and Vincze (1978) that

$$e^{t-c\sqrt{t}} < f(t) < e^{t+c\sqrt{t}}$$

with some constant $c > 0$. Based on this result, a final answer (i.e., under the given conditions, it follows that $f(t) = e^t$) was given by Miles and Williamson (1986).

**Acknowledgments**

The author is grateful to the referees for their valuable comments.

**References**


