The Annals of Statistics
1975, Vol. 3, No. 5, 1031–1037

JAROSLAV HÁJEK, 1926–1974

BY VÁCLAV DUPAČ

Charles University, Prague

Jaroslav Hájek died in Prague on June 10, 1974, at the age of 48. He worked till his last days, although suffering severely from a longlasting kidney disease. His untimely death is a heavy loss for mathematical statistics as a science as well as for his students and co-workers all over the world. Hájek's results and ideas have significantly influenced the development of many aspects of mathematical statistics since the late fifties and they continue to stimulate further research and progress.

Jaroslav Hájek was born in Poděbrady, Czechoslovakia, on February 4, 1926. He attended the Technical University in Prague from 1945 to 1949, leaving with the degree of a statistical engineer. After a postgraduate fellowship under the guidance of J. Novák, he received the C. Sc. degree (equivalent to the Ph. D.) in 1955. From 1954 to 1964 he worked as a researcher at the Mathematical Institute of the Czechoslovak Academy of Sciences. In 1964 Hájek was appointed Head of the Department of Mathematical Statistics at Charles University in Prague, with appointment to Professor in 1966. He was a visiting professor at the University of California at Berkeley in the academic year 1965–66 and at the Florida State University in 1969–70. Hájek was a Fellow of the Institute of Mathematical Statistics, Associate Editor of the Annals of (Mathematical) Statistics from 1970, and member of the editorial board of four other international journals. For his outstanding contribution to mathematical statistics, he was awarded the Klement Gottwald National Prize in 1973.

Hájek's contribution to statistics (and probability) is of great scope. It includes research in the theory of rank tests, parametric estimation, probability sampling, statistical inference in stochastic processes and various other specializations. The striking feature of Hájek's papers is the originality of his ideas: results were usually achieved by developing new methods of proof, which then provided useful tools for the solution of many related problems.

In a series of papers, later included (unified and complemented) in the monograph (1967: 4), Hájek investigated properties of linear rank statistics and tests based on them. Linear rank statistics are of the form $S = \sum_{i=1}^{N} c_i a(R_i)$, where the R_i are the ranks of independent random variables X_i with distribution functions F_i, the c_i are regression constants, and the $a(i)$ are rank scores, all symbols depending on N in general. Hájek (1961: 27) gave necessary and sufficient conditions of the Lindeberg type for the asymptotic normality of S under the null

---

Received November 1974; revised March 1975.

AMS 1970 *subject classifications*. Primary 01A70; Secondary 00A15.

Key words and phrases. Obituary, nonparametric statistics, asymptotic theory, rejective sampling, estimation in stochastic processes.

hypothesis $F_1 = \cdots = F_N$. If both the a_i's and c_i's are infinitesimal when normalized to zero sums and unit sums of squares, then the condition reads $\lim_{N\to\infty} \sum\sum_{|c_i a_j|>\varepsilon} c_i^2 a_j^2 = 0$ for every $\varepsilon > 0$. Making use of Le Cam's concept of contiguity, Hájek proved the asymptotic normality of S under contiguous alternatives, restricting consideration to absolutely continuous underlying distributions with finite Fisher information. For example, under location alternatives with likelihood function $\prod_{i=1}^N f(x_i - d_i)$ compared with $\prod_{i=1}^N f(x_i)$, the conditions $\max_{1\leq i\leq N} (d_i - \bar{d})^2 \to 0$ and $I(f) \sum_{i=1}^N (d_i - \bar{d})^2 \to b$, $0 < b < +\infty$ establish the asymptotic normality of S with simple expressions for asymptotic mean and variance. This result enables one to find asymptotically most powerful rank tests against fixed contiguous alternatives. Similar results were derived for scale alternatives, for alternatives to the hypothesis of symmetry and for the k-sample case with χ^2-type statistics. Extension of the Kolmogorov–Smirnov test to regression alternatives and a derivation of the limit distribution of the test statistic are also due to Hàjek (1965: 35). Later, Hájek (1968: 39; 1969: 40) gave a far-reaching generalization of the Chernoff–Savage theorem concerning asymptotic normality of S under noncontiguous alternatives. The projection method developed here, together with the ingenious inequality, $\operatorname{Var} \sum_{i=1}^N c_i a(R_i) \leqq 21 \max_{1\leq i\leq N} (c_i - \bar{c})^2 \sum_{i=1}^N [a(i) - \bar{a}]^2$, have found wide application in further research. In this last paper on rank tests (1974: 46), Hájek proved asymptotic sufficiency of the vector of ranks in the Bahadur sense for testing randomness against a general class of two-sample alternatives.

Although his results in rank test theory are probably the best known ones, Hájek's beloved topic was probability sampling from finite populations. He established (1960: 25) necessary and sufficient conditions for asymptotic normality of estimators based on simple random sampling without replacement. The solution of this problem was obtained by approximating simple random sampling by Poisson (binomial) sampling, which may be decomposed into independent subexperiments and is thus easier to study. A refinement of this method was used (1964: 34) for the study of rejective sampling. Rejective sampling consists of selection of n items with replacement according to probabilities α_i; if the items are not distinct, the sample is rejected and a new sample of n items is selected. Poisson (binomial) sampling is defined as independent selection of the ith item with probability p_i. Relating the p_i's to α_i's in a proper way, Hájek obtained approximations for rejective sampling to π_i and π_{ij}, the probabilities of inclusion of item i and both items i and j in the sample, and to the variance of the unbiased linear estimator $\sum_{i\in\text{sample}} y_i/\pi_i$ of $\sum_{i=1}^N y_i$, and necessary and sufficient conditions for asymptotic normality of a modified estimator. Hájek has also contributed to the theory of ratio estimators (1958: 20) and to the problem of selection of optimal sampling designs, balancing the accuracy of estimates and the cost of experimentation (1959: 22). Hájek (1960: 1) published a monograph on probability sampling in Czech and he left an almost completed manuscript of an entirely new book on the same topic.

In the theory of parametric estimation, Hájek (1970: 43) proved, under the assumption of local asymptotic normality, that the limiting distribution of the estimator of a k-dimensional parameter is a convolution of a certain normal distribution, which depends only on the underlying distributions, and of another distribution, which depends on the choice of the estimator. Under the same assumption and for any sequence of estimates $\{T_n\}$ of θ_1 (with θ_j, $2 \leqq j \leqq k$, as nuisance parameters) and for a general class of loss functions l, Hájek found a simple lower bound for the local asymptotic minimax risk

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \sup_{|\theta - t|<\delta} E_\theta\{l(n^{\frac{1}{2}}(T_n - \theta_1))\},$$

as well as conditions for the lower bound to be attained (1972: 45).

In the period 1955–1961, Hájek contributed significantly to statistical inference in stochastic processes. One of his early results (1956: 12) follows: Let X_t be a stationary process with mean m, variance σ^2 and convex (normed) correlation function $R(\tau)$. If $\hat{m}$ is a linear estimator for m of the type $\hat{m} = \int_0^T X_t \, d\Phi(t)$, Φ of bounded variation, $\Phi(0) = 0$, $\Phi(T) = 1$, then $\liminf_{T\to\infty} T \operatorname{Var} \hat{m} \geqq 2\sigma^2 \int_0^\infty R(\tau)\, d\tau$. In the article (1961: 28) he showed that classical theorems on least-square estimation of $\theta = \sum_{i=1}^m c_i \alpha_i$ remain true for any infinite family $\{X_t\}$ with any covariance structure and any functions φ_{it} involved in the linear hypothesis $E_\alpha X_t = \sum_{i=1}^m \alpha_i \varphi_{it}$. A unified theoretical approach to problems of linear statistical inference on stochastic processes was developed in the paper (1962: 30) together with explicit results for some particular classes of stationary processes.

Only one of Hájek's papers can be counted in pure probability theory. However, his early result, a slight generalization of which is known as the Hájek–Rényi inequality (1955: 9), has entered standard textbooks on probability theory:

$$P(\sup_{k\geqq n} k^{-1}|X_1 + \cdots + X_k| \geqq \varepsilon) \leqq \varepsilon^{-2}(n^{-2} \textstyle\sum_{k=1}^n \operatorname{Var} X_k + \sum_{k=n+1}^\infty k^{-2} \operatorname{Var} X_k)$$

for independent random variables X_k with zero means and finite variances.

THE PUBLICATIONS OF JAROSLAV HÁJEK

BOOKS

1960

[1] *Theory of Probability Sampling with Applications to Sample Surveys.* Nakladatelství ČSAV, Prague (in Czech).

1962

[2] *Probability in Science and Engineering* (with V. Dupač). Nakladatelství ČSAV, Prague (in Czech).

1967

[3] *Probability in Science and Engineering* (with V. Dupač). Academia, Prague and Academic Press, New York.

[4] *Theory of Rank Tests* (with Z. Šidák). Academia, Prague and Academic Press, New York. (Russian translation: Izd. Nauka, Moscow 1971.)

1969

[5] *A Course in Nonparametric Statistics.* Holden-Day, San Francisco.

ARTICLES

1949

[6] The use of factorial design and of confidence intervals in weighing. *Statistický Obzor* **29** 258–273 (in Czech).

[7] Representative sampling by a two-stage method. *Statistický Obzor* **29** 384–394 (in Czech).

1955

[8] Some rank distributions and their applications. *Časopis Pěst. Mat.* **80** 17–31 (in Czech). (English translation in *Selected Transl. Math. Statist. Prob.* **2** (1962) 27–40.)

[9] Generalization of an inequality of Kolmogorov (with A. Rényi). *Acta Math. Acad. Sci. Hungar.* **6** 281–283.

[10] On some fundamental questions of mathematical statistics (with F. Fabian). *Časopis Pěst. Mat.* **80** 387–399 (in Czech).

1956

[11] Asymptotic efficiency of a certain sequence of tests. *Czechoslovak Math. J.* **6** 26–30 (in Russian).

[12] Linear estimation of the mean value of a stationary random process with convex correlation function. *Czechoslovak Math. J.* **6** 94–117 (in Czech). (English translation in *Selected Transl. Math. Statist. Prob.* **2** (1962) 41–61.)

[13] Stratified sampling. *Apl. Mat.* **1** 149–161 (in Czech).

[14] Remark on the article "On certain sequences of sets of points on a circle". *Časopis Pěst. Mat.* **81** 77–78 (in Czech).

1957

[15] Inequalities for the generalized student's distribution and their applications. *Časopis Pěst. Mat.* **82** 182–194 (in Czech). (English translation in *Selected Transl. Math. Statist. Prob.* **2** (1962) 63–74.

1958

[16] Predicting a stationary process when the correlation function is convex. *Czechoslovak Math. J.* **8** 150–154.

[17] A property of *J*-divergence of marginal probability distributions. *Czechoslovak Math. J.* **8** 460–463.

[18] On a property of normal distributions of any stochastic process. *Czechoslovak Math. J.* **8** 610–618 (in Russian). (English translation in *Selected Transl. Math. Statist. Prob.* **1** (1961) 245–252.)

[19] On the distribution of some statistics in the presence of intraclass correlation. *Časopis Pěst. Mat.* **83** 327–329 (in Czech). (English translation in *Selected Trasl. Math. Statist. Prob.* **2** (1962) 75–77.)

[20] On the theory of ratio estimates. *Apl. Mat.* **3** 384–398.

[21] Some contributions to the theory of probability sampling. *Bull. Inst. Internat. Statist.* **36** No. 3 127–134.

1959

[22] Optimum strategy and other problems in probability sampling. *Časopis Pěst. Mat.* **84** 387–423.

[23] The age of eruption of permanent teeth in children in Czechoslovakia (with V. Poncová). *Ceskoslovenská Stomatologie* **2** 104–113 (in Czech).

1960

[24] On a simple linear model in Gaussian processes. *Transactions of the 2nd Prague Conference on Information Theory, Statistical Decision Functions, Random Processes.* Liblice, 1959 (J. Kožešník, ed.) 185–197.

[25] Limiting distributions in simple random sampling from a finite population. *Publ. Math. Ints. Hungar. Acad. Sci.* **5** 361–374.

1961

[26] On plane sampling and related geometrical problems (with T. Dalenius and S. Zubrzycki). *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 125–150. Univ. of California Press.

[27] Some extensions of the Wald–Wolfowitz–Noether theorem. *Ann. Math. Statist.* **32** 506–523.

[28] On linear estimation theory for an infinite number of observations. *Teor. Verojatnost. i Primenen.* **6** 182–193.

[29] Concerning relative accuracy of stratified and systematic sampling in a plane. *Colloq. Math.* **8** 133–134.

1962

[30] On linear statistical problems in stochastic processes. *Czechoslovak Math. J.* **12** 404–444.

[31] An inequality concerning random linear functionals on a linear space with a random norm and its statistical applications. *Czechoslovak Math. J.* **12** 486–491.

[32] Asymptotically most powerful rank-order tests. *Ann. Math. Statist.* **33** 1124–1147.

[33] Cost minimization in multiparameter estimation. *Apl. Mat.* **7** 405–425 (in Czech).

1964

[34] Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist.* **35** 1491–1523.

1965

[35] Extension of the Kolmogorov–Smirnov test to regression alternatives. *Bernoulli–Bayes–Laplace*; Proceedings of an International Research Seminar, 1963 (J. Neyman and L. LeCam, eds.) 45–60 Springer-Verlag, Berlin.

1967

[36] On basic concepts of statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 139–162. Univ. of California Press.

1968

[37] Locally most powerful rank tests of independence. *Studies in Math. Statist.* (K. Sarkadi and I. Vincze, eds.) 45–51. Akádémiai Kiadó, Budapest.

[38] Some new results in the theory of rank tests. *Studies in Math. Statist.* (K. Sarkadi and I. Vincze, eds.) 45–51. Akadémiai Kiadó, Budapest.

[39] Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

1969

[40] Asymptotic normality of simple linear rank statistics under alternatives II (with V. Dupač). *Ann. Math. Statist.* **40** 1992–2017.

[41] Asymptotic normality of the Wilcoxon statistic under divergent alternative (with V. Dupač). *Zastos. Mat.* **10** 171–178.

1970

[42] Miscellaneous problems of rank test theory. *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 3–19. Cambridge Univ. Press.

[43] A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 323–330.

1971

[44] Limiting properties of likelihood and inference. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.). Holt, Rinehart and Winston, Toronto.

1972

[45] Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175–194. Univ. of California Press.

1974

[46] Asymptotic sufficiency of the vector of ranks in the Bahadur sense. *Ann. Statist.* **2** 75–83.

[47] Regression designs in autoregressive stochastic processes (with G. Kimeldorf). *Ann. Statist.* **2** 520–527.

[48] Asymptotic theories of sampling with varying probabilities without replacement. *Proc. Prague Symp. on Asymptotic Statistics* (1973) (J. Hájek, ed.) **1** 127–138.

This list does not include chapters on probability and statistics in collegiate mathematics textbooks or in mathematical manuals, lecture notes for students, preliminary communications and abstracts.

DEPARTMENT OF MATHEMATICAL STATISTICS
CHARLES UNIVERSITY
SOKOLOVSKA 83
18600 PRAGUE 8
CZECHOSLAVAKIA