JAROSLAV HÁJEK

Jan Coufal – Petra Dotlačilová – Jiří Tobíšek

Abstract

Despite his relatively short life, Jaroslav Hájek probably became the most prominent Czech theoretical statistician. Hájek studied statistical and insurance engineering at the College of Special Sciences of the Czech Technical University and he successfully completed this study by obtaining an engineering degree in 1949. In 1947, although only two years into his undergraduate course, Hájek began teaching as an Assistant Lecturer in the Mathematical Institute at the Czech Technical University where he studied. In 1952, he was free to begin his postgraduate research studies which he undertook at the Mathematical Institute of the Czechoslovak Academy of Sciences. In 1954, he received the title of CSc. for the paper Contributions to the theory of statistical estimation. Hájek continued to work in the Mathematical Institute as a Senior Research Worker. In 1963, he received a DrSc. and still in the same year he received his habilitation at the Faculty of Mathematics and Physics of Charles University. In 1966, he was entitled professor at this faculty. In 1973, he was awarded the State Prize for his work on the asymptotic theory of ordinal tests. He died at the age of 48.

Key words: Jaroslav Hájek, Charles University, Czech Technical University, mathematical statistics, asymptotic theory of ordinal tests

JEL Code: B26, G22

Introduction

The Johnson et al. (1997) contains the biographies of the 114 most important (according to the book's authors) world statisticians from the 17th century to 1997. Mathematicians involved in probability theory were also included in the book, but no living (at the time of the book's publication) people were included. It is understandable that any such review is always subjective to a certain extent, however, since the mentioned book was published by a renowned publishing house that systematically publishes statistical literature, the selection of biographies in this book can be considered a fairly accurate picture of the current view of the history of probability theory and mathematical statistics on a global scale. Jaroslav Hájek is the only one mentioned in this book among the mathematicians working in the Czechia, who,

according to the mentioned book, can be considered the most important Czech statistician in the history of Czech mathematics (except for mathematicians living in 1997 and later).

1 Biography

According to Šidák (1975, 1976), Jaroslav Hájek was born on February 4, 1926 in Poděbrady (Czechoslovakia) in a simple family — his father was a barber, his mother a seamstress. Soon he lost his father, and the material conditions of mother and the little boy became rather cramped. In these years and later, Jaroslav's father was replaced by his uncle Martínek, a painter from Poděbrady. Although his mother remarried about five years later, she was widowed again in 1944. After all, despite insufficient financial resources, she allowed her son to study, since he showed great mathematical talent from a young age. After graduating from the Prague grammar school, Hájek studied statistical and insurance engineering at the College of Special Sciences at the Czech University of Technology (hereafter CTU), where he graduated in June 1949 as an engineer (title Ing.), prof. Janko and prof. Schoenbaum lectured here (Coufal et al., 2021). At this time he also published his first professional article (Hájek, 1949). From 1948 to 1950, he worked as an assistant at the Department of Mathematics at CTU, teaching at the College of Special Sciences CTU. In 1950, he defended his rigorous thesis and was awarded a doctorate in technical sciences (title RTDr.). Hájek's secondary school education had been disrupted by the war but this was not the end of the disruption to his career for after receiving his Diploma he had to spend two years doing military service. Then, in the years 1952 - 54, he was an aspirant at the Mathematics Institute of the Czechoslovak Academy of Sciences (hereafter ČSAV), where in 1955 he defended his candidate's dissertation Contributions to the Theory of Statistical Estimation (supervisor of this thesis was J. Novák) and received the degree of CSc. From 1954, he worked as a researcher at this institute until 1966. In 1963, he became a doctor of physical and mathematical sciences (title DrSc.) on the basis of a dissertation on statistical problems in stochastic processes. In 1955, he married Alžběta Galambosová, and happy years followed, when two daughters were born to the couple one after the other. He began his teaching activities in 1948-50 as an assistant at the CTU. He also lectured at the University of Economics and Business in Prague and at the Faculty of Mathematics and Physics of Charles University (hereafter MFF UK). In 1963, he completed his habilitation at the MFF UK and began lecturing there. In 1964, he became the external head of the Department of Mathematical Statistics at the MFF UK and then headed this department for many years.

Finally, in 1966, he was appointed a professor at the MFK UK and moved here permanently. In 1973, he received one of the highest recognitions for his scientific work – he was awarded the state prize for developing the asymptotic theory of statistical ordinal tests. Prof. Hájek died in Prague on 10 June 1974 at the age of only 48 after a kidney transplant (Johnson et al., 1997). His death meant an irreplaceable loss for mathematical statistics on a global scale, because Hájek was the founder and foremost representative of the modern scientific mathematical-statistical school in the post-war period.

2 Hájek's scientific activity

According to Šidák (1975, 1976) prof. Hájek also performed a number of functions in the organization of Czechoslovak and international scientific life with a strong interest in scientific progress. He was a member of the scientific council of the MFF UK and chairman of the commissions for the defense of candidate and doctoral dissertations in the field of probability theory and mathematical statistics, a member of the scientific college of mathematics and a member of the commission for the use of mathematical methods in economics at the Institute of Economics of the ČSAV. He served on the editorial boards of the following international journals: Annals of Mathematical Statistics, Advances in Applied Probability, Zeitschrift fiir Wahrscheinlichkeitstheorie und verwandte Gebiete, Mathematische Operationsforschung und Statistik, Czechoslovak Mathematical Journal, formerly also Applied Mathematics. From international scientific organizations, he was an honorary member (Fellow) of the Institute of Mathematical Statistics, a member of the International Statistical Institute and the Econometric Society. Thanks to his scientific reputation, Hájek was invited many times abroad for working stays and lectures at conferences; he often had so many invitations that it was not even possible for him to accommodate them all. For all of them, we will only mention his four longer stays in the USA, where he was invited to Berkeley at the University of California (three times), to East Lansing at Michigan State University and to Tallahassee at Florida State University; to clarify, let's say that, for example, during his last stay in the USA (he received an invitation to a lecture from 45 universities on the American continent). In his scientific work, prof. Jaroslav Hájek dealt mostly with the areas of mathematical statistics – theory of sample surveys, theory of ordinal tests and statistical problems in stochastic processes. A typical feature of his work was that he never denied his original engineering training (in the best sense of the word) and saw to the real nature of problems. Such an approach with an

understanding of the needs of practice combined with deep, sophisticated and modern mathematical methods were the source of many of his significant and stimulating results. At the time of their creation, his publications always advanced the development of the relevant field by a significant amount and were followed up countless times by foreign and our authors. Several fundamental results are often cited under Hájek's name (Strutz, 2016). Among Hájek's favorite areas, the theory of sample surveys was the first (Hájek, 1949a, 1955); he remained faithful to her until the end of his life (Hájek, 1974a, 1969). In Hájek (1955) the estimates are in the regional (stratified) sample and the optimal, resp. by randomly distributing the selection into areas. Hájek (1948b) contributes to the asymptotic theory of ratio estimates by using the regression method and the method based on the normal distribution. In Hájek (1959), sampling plans and estimation methods are found, which optimally solve the conflict between the costs of the experiment and the accuracy of the estimates in the Bayesian approach; this work touches on Hájek's third interest — stochastic processes, as the solution is given not only for uncorrelated random variables, but also for stationary random sequences with a convex correlation function (Bellingsley, 2013). Another optimization problem is solved in Hájek (1962a), namely when estimating multiple parameters. Hájek (1960) states a necessary and sufficient condition for the asymptotic normality (or also for the asymptotic Poisson distribution) of estimates in simple random sampling without returns; the proof method is based on the approximation of simple random sampling by Poisson (binomial) sampling, which is defined as independent selection of elements with certain probabilities, so it is easier to handle theoretically. This idea is followed by an article by Hájek (1964) on rejection selection, i.e. a selection in which n elements are independently selected with generally different probabilities and returns, but if all selected elements are not different from each other, the entire selection is rejected and taken new selection; here again the Poisson sampling approximation is used to obtain results on asymptotic normality. The articles by Hájek (1961) and Hájek et al. are devoted to a completely different issue, important, for example, in geological research — the selection of points in a plane. (1961). Hájek published the book Hájek (1960a) on the theory of sample surveys, the first part of which is essentially a basic textbook, the second part a monograph, summarizing some of his original results; furthermore, not long before his death, he completed the manuscript of the new monograph Hájek (1969), which he, however, intended to revise again. The second area of Hájek's intense interest was the theory of ordinal tests, especially their asymptotic theory. His works here literally marked milestones in the development of this theory, and Hájek was considered one of the leading experts – or perhaps

even the first expert - in this field in the entire world. Already at a time when the first articles about sequential tests were just beginning to appear timidly, Hájek sensed their future importance and dealt with them in his dissertation submitted in 1949; in its part, later published in Hájek (1955), he derived generating functions and proved the asymptotic normality of the distribution of statistics, which are now known under the names Wilcoxon two-sample and one-sample statistics and Kendall's rank correlation coefficient. In Hájek et al. (1955) necessary and sufficient Lindeberg-type conditions for asymptotic normality under the null hypothesis are found. The relevant sentence is now quoted by Hájek's name, or under the names Wald-Wolfowitz-Noether-Hájek. Hájek (1962) is a significant contribution to the use of the concept of contiguity in the theory of ordinal tests, which was originally introduced by LeCam for other purposes. Research in this line later continued with the works of Hájek (1968) and Hájek et al. (1969), where asymptotic normality is proven for very general non-contiguous regression alternatives, and similar even stronger results for the Wilcoxon statistic in Hájek (1969a); these are far-reaching generalizations of the well-known Chernoff-Savage theorem, and the ingenious proof method is based on a remarkable new inequality for variances and an approximation using its projections onto sums of independent quantities. In Hájek (1970), in addition to several other problems, he mainly dealt with density estimation for the purpose of choosing suitable scores in ordinal statistics. In Hájek (1974) he investigated efficiency in Bahadur's sense and proved that when testing randomness against a two-choice alternative, ordinal tests achieve the best possible exact guidelines. Hájek's serial test results up to about 1965 were then summarized, systematized and supplemented with a number of other results in a three-hundred-page monograph by Hájek et al. (1967); in addition to the mentioned statistics for testing two samples and regression, ordinal statistics for testing symmetry in one sample, statistics for testing independence, and statistics of the Kolmogorov-Smirnov and Cramérov-von Mises types are similarly investigated. After this thorough scientific work, Hájek (1969) intended another book from this area as a textbook for university students and therefore wrote it at a more accessible level. The third broader area of the scientific work of prof. Hájek was statistical problems in stochastic processes. Here he first paid attention to stationary processes with a convex correlation function: he found a lower bound for the variance of linear estimates of the mean of such processes and showed that the variance of the usual estimate using the mean just reaches this lower bound; similarly, Hájek (1958) found a lower bound for the residual variance of linear prediction in these processes. Based on the limiting properties of J-divergences from Hájek (1958a), it was proved that the probability measures of any two Gaussian processes are either equivalent or

mutually singular; the same result was published in the same year by J. Feldman and is therefore now cited as the Feldman-Hájek's dichotomy theorem, but Hájek's proof method is more constructive and therefore more useful for investigating special cases. The long article by Hájek (1962) is again of fundamental importance, as it develops a unified theoretical approach to solving a number of problems such as prediction, filtering, estimation of regression parameters, etc. in stochastic processes, based on the correspondence of linear subspaces created on the one hand by random quantities xř, on the other hand by their covariances (around the same time or a little later, other authors developed a similar approach to these problems based on the so-called Hilbert spaces reproducing the kernel.). Furthermore, Hájek (1962) analyzes the conditions under which the solution of these problems can be interpreted using individual trajectories (i.e. not only as limits according to the center), the strong equivalence of the distribution of Gaussian processes is defined and the density of such a distribution relative to another is studied; various general results of the paper are explicitly elaborated for stationary processes with rational spectral density (Strock, 2013). In addition to the discussed three areas of long-term and intensive interest of prof. Hájek also contributed to some other areas. Although these are rather individual articles, many of them also contain very significant results; we will only briefly mention some of them. Thus, for example, in Hájek et al. (1955), an interesting inequality (a generalization of Kolmogorov's inequality) for the probabilities of the sums of independent quantities is proved, which has been used, generalized and quoted many times in textbooks and which is now generally known as the Hájek‘s-Rényi inequality (Rényi, 1972). Next, we present two contributions to the parametric theory of estimation: in Hájek et al. (1970a) is proved the remarkable result that in a wide class of cases the asymptotic distribution of estimates is a convolution of a certain normal distribution that depends only on the underlying distribution and another distribution that depends only on the choice of estimate; under similar assumptions, Hájek (1972) found a lower bound for the local asymptotic minimax risk of estimates for a very general loss function. Hájek had an admirable gift of a kind of engineering insight into the essence of statistical problems. Therefore, he was keenly interested in various basic questions of statistical induction. He often discussed these problems with interest, mentions of them can be found in some of his publications, but unfortunately he devoted only three articles to them (Hájek 1955, 1966 and 1967). The last of them, Hájek's lecture at the Berkeley symposium, is the most significant and analyzes the concepts of sufficiency, invariance, similarity, conditionality and plausibility.

3 Hájek's publishing activity

Only a selection of both articles and books is listed. A detailed inventory is in Šidák (1975):

HÁJEK, J. (1949). Užití komplexní methody a intervalu spolehlivosti při vážení. *Statistický obzor*, Vol 29, pp. 258-273.

HÁJEK, J. (1949a). Representativní výběr skupin methodou dvou fází. *Statistický obzor*, Vol. 29, pp. 384-394.

HÁJEK, J. (1955). Některá pořadová rozdělení a jejich použití. *Čas. pěst. mat.* Vol. 80, pp. 17-31.

HÁJEK, J. (1955a). *Teorie výběrových šetření*. Praha : SPN.

HÁJEK, J. (1956). Oblastní vyběr. *Aplikace matematiky*. Vol. 1, pp. 149-161.

HÁJEK, J. (1958). Predicting a stationary process when the correlation function is convex. *Czechoslovak Math. J.* Vol. 8, pp. 150-154.

HÁJEK, J. (1958a). A property of J-divergences of marginal probability distributions. *Czechoslovak Math. J.* Vol. 8, pp. 460-463.

HÁJEK, J. (1958b). On the theory of ratio estimates. *Aplikace matematiky* 3, pp. 384—398.

HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Čas. pěst. mat*. Vol. 84, pp. 387-423.

HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publ. Math. Inst. Hung. Acad. Sci.* Vol. 5, pp. 361-374.

HÁJEK, J. (1960a). *Teorie pravděpodobnostního výběru s aplikacemi na výběrová šetření*. Praha : NČSAV.

HÁJEK, J. (1961). Concerning relative accuracy of stratified and systematic sampling in a plane. *Colloquium math*. Vol. 8, pp. 133-134.

HÁJEK, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Statist*. Vol. 33, pp. 1124—1147.

HÁJEK, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist*. Vol. 35, pp. 1491-1523.

HÁJEK, J. (1966). On the foundations of statistics. *Conf. on Math. Methods in Economic Research, Smolenice 1965*. Praha : Ekonomicko-matematická laboratoř při Ekonomickém ústavu ČSAV, pp. 1-17

HÁJEK, J. (1967). On basic concepts of statistics. *Proc. 5th Berkeley Symp. Math. Statist. Prob. 1965—1966.* Berkeley : Univ. of Calif. Press, vol. I, pp. 139-162.

HÁJEK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives, *Ann. Math. Statist*. Vol. 39, pp. 325-346.

HÁJEK, J. (1969). *A Course in Nonparametric Statistics*. San Francisco : Holden-Day.

HÁJEK, J. (1970). Miscellaneous problems of rank test theory. *Nonparametric Techniques in Statistical Inference* (ed. M. L. Puri). Cambridge : Cambridge Univ. Press, pp. 3-17.

HÁJEK, J. (1970a). A characterization of limiting distributions of regular estimates. *Zeit. Wahrscheinlichkeitstheorie verw*. Vol. I4, pp. 323-330.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. 6th Berkeley Symp. Math. Statist. Prob. 1970*. Berkeley : Univ. of Calif. Press. Vol. I, pp. 175-194.

HÁJEK, J. (1974). Asymptotic sufficiency of the vector of ranks in the Bahadur sense. *Ann. Statist.* Vol. 2, pp. 75-83.

HÁJEK, J. (1974a). Asymptotic theories of sampling with varying probabilities without replacement. *Proc. Prague Symp. on Asymptotic Statistics, September 1973*, Praha : Universita Karlova, Vol. I, pp. 127-138.

HÁJEK, J., DALENIUS, T., & ZABRZYCKI, D. (1961). On plane sampling and related geometrical problems. *Proc. 4th Berkeley Symp. Math. Statist. Prob. 1960,* Berkeley : Univ. of Calif. Press. Vol. I, pp. 125-150.

HÁJEK, J., & DUPAČ, V. (1969). Asymptotic normality of the Wilcoxon statistic under divergent alternatives. *Zastosowania matematyki*. Vol. I0, pp. 171-178.

HÁJEK, J., & DUPAČ, V. (1969a). Asymptotic normality of simple linear rank statistics under alternatives II. *Ann. Math. Statist*. Vol. 40, pp. 1992-2017.

HÁJEK, J., & FABIÁN, F. (1955). K některým základním otázkám matematické statistiky. *Čas. pěst. mat.* Vol. 80, pp. 387-399.

HÁJEK, J., & RÉNYI, A. (1955). Generalization of an inequality of Kolmogorov. *Acta Math. Acad. Sci. Hung*. Vol. 6, pp. 281-283.

HÁJEK, J., & ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Praha : Academia, , & New York : Academia Press, New York.

Conclusion

However, his scientific achievements and positive attitude to life were fought hard against fate: he suffered from serious kidney disease for a large part of his life. Especially in the last years and months, it was admirable how bravely and with extraordinary willpower he fought

his illness. Despite his deterioration, he still worked scientifically and organizationally and retained his vitality and understanding for the needs of others. Even in the last weeks of his life, he had many plans for further scientific work, which unfortunately remained unfulfilled. According to David et al. (2001) in the person of prof. Hájek, the most important Czech and world-renowned expert in mathematical statistics, has passed away. Words about the irreplaceability of this loss are not just a phrase here, because for a really long time there will be no one in our country who would be able to take a place in scientific life in the full sense of the word such outstanding personalities according to Anděl (2019), Zvára et al. (2019), Antoch et al. (2022). One of the authors, whose mother studied statistics and insurance engineering at the College of Special Sciences of CTU with J. Hájek in 1945-1949, would like this article to be a repayment of the debt to prof. Hájek and also served as thanks for his advice during the same author's studies at the MFF UK, which he gave him.

Acknowledgment

This contribution is a follow-up to the project of University of Economics and Management Prague.

References

ANDĚL, J. (2019). *Statistické metody*. Praha : Matfyzpress. ISBN: 978-80-7378-381-5.

ANTOCH, J., MOLA, F., & VOZÁR, O. (2022). New Randomized Response Technique for Estimating the Population Total of a Quantitative Variable. *Statistika* : *Statistics and Economy Journal*, Vol, 102. No 2, pp. 205-227. URL: https://doi.org/10.54694/stat.2022.11.

BILLINGSLEY, P. (2013). *Convergence of probability measures*. New York: John Wiley & Sons.

COUFAL, J., DOTLAČILOVÁ, P., & TOBÍŠEK, J. (2021). Emil Schoenbaum. *The 15th International Days of Statistics and Economics, September 9 – 11, 2021*, Prague, Czech Republic – Conference Proceedings, Slaný : Macáková, Melandrium, pp. 173-182, URL: https://msed.vse.cz/msed_2021/sbornik/toc.html, ISBN 978-80-87990-25-4.

DAVID, H. A., & EDWARDS, A. W. F. (2001). *Annotated Readings in the History of Statistics*. New York : Springer, pp. 7–18. ISBN: 978-0387988443.

JOHNSON, N. L., & KOTZ, S. (1997). *Leading Personalities in Statistical Sciences*: *From the Seventeenth Century to the Present*. New York etc. : J. Wiley & Sons, pp.194 – 196, ISBN 0-471-16381-3.

RÉNYI, A. (1972). *Teorie pravděpodobnosti*. Praha : Academia.

STROOCK, D. W. (2013). *An introduction to Markov processes*. Berlin etc. : Springer. ISBN 3-540-23451-9

STRUTZ, T. (2016). *Data Fitting and Uncertainty* (*A Practical Introduction to Weighted Least Squares and Beyond*) (2nd ed.). Wiesbaden : Vieweg & Teubner. ISBN 978-3-658-11455-8.

ŠIDÁK, Z. (1975). Zemřel professor Jaroslav Hájek. *Casopis pro pěstování matematiky*. Vol. 100, No. 3, pp. 303-313. URL: http://dml.cz/dmlcz/117881.

ŠIDÁK, Z. (1976). Zemřel professor Jaroslav Hájek. *Aplikace matematiky*. Vol. 21, No. 1, pp. 77-80. URL: http://dml.cz/dmlcz/103623.

ZVÁRA, K., & ŠTĚPÁN, J. (2019). *Pravděpodobnost a matematická statistika*. Praha : Matfyzlpress. ISBN: 978-80-7378-388-4.

Contact

Jan Coufal
University of Economics and Management
Nárožní 2600/9a, 158 00 Prague 5, Czech Republic
jan.coufal@vsem.cz

Petra Dotlačilová
Prague University of Economics and Business
Faculty of Informatics and Statistics (Dep. of Mathematics)
W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic
petra.dotlacilova@vse.cz

Jiří Tobíšek
University of Economics and Management
Nárožní 2600/9a, 158 00 Prague 5, Czech Republic
jiri.tobisek@vsem.cz