CWI Quarterly Volume 7 (3) 1994, pp. 259 – 266

Minisymposium in Honour of Jaroslav Hájek 8th June 1994 Charles University, Prague, Czech Republic

Roelof Helmers
CWI
P.O. Box 94079 1090 GB Amsterdam The Netherlands

A report is given on the Minisymposium held in honour of Jaroslav Hájek on 8th June 1994 in Prague.

1. Introduction

Twenty years ago Jaroslav Hájek died in Prague on the 10th June 1974 at the early age of 48. To commemorate Hájek and his work, The Department of Probability and Mathematical Statistics, Charles University, Prague, organized an one-day Symposium on 8th June 1994. This meeting brought together about thirty researchers working in various areas of mathematical statistics, where the significant impact of Hájek's results and ideas is still present even today: the theory of rank tests, efficiency problems in estimation theory, sampling from a finite population, and a few miscellaneous topics. Václav Dupač (Prague) gave the opening talk on Jaroslav Hájek and the influence of his work on Czech statistics (see also [1]); Hájek's widow and two daughters also attended the meeting.

In this report I shall give a brief overview of the talks, which at the same time will emphazise the lasting importance of Hájek's work in mathematical statistics. At this point it should also be mentioned that Hájek initiated the Prague Symposia on Asymptotic Statistics. The first of these meetings was held in August 1973. Since then, these symposia have been held each five years and have greatly stimulated contacts between Czech mathematical statisticians and their colleagues from abroad. This was especially important before the 'Velvet Revolution' took place in November 1989.

A forthcoming issue of *Kybernetika* will contain the invited papers of the Minisymposium as well as a number of contributed papers.

2. Theory of rank tests

The heading of this section is also the title of the famous research monograph by Hájek and Šidák ([6]) which was published in 1967. Many graduate students at the time (including myself) learned about contiguity, the three Lemmas of LeCam, and many results on rank tests from this beautiful book. The book by Witting and Nölle [17] was another very good source on these matters.

At the meeting Jana Jurečková (Prague) and H. Witting (Freiburg) gave invited lectures on different topics in this area. Jana Jurečková gave a survey of Hájek's asymptotic theory of linear rank tests and also discussed some recent extensions (see, e.g., Guttenbrunner and Jurečková [3]) of Hájek's results to the linear regression model. To give the flavour of some of Hájek's ideas in this area we need first to introduce a bit of notation.

Let $X_1, \ldots, X_n$ be independent observations, where X_i has continuous distribution function (df) F_i $(1 \leq i \leq n)$. Let R_{ni} denote the *rank* of X_i among $X_1, \ldots, X_n$; i.e.

$$R_{ni} = \sum_{j=1}^{n} I_{[o,\infty)}(X_i - X_j), \ 1 \leq i \leq n \tag{1}$$

where $I_{[0,\infty)}$ denotes the indicator function of the set $[0, \infty)$. The rank of X_i is nothing but the number of observations among $X_1, \ldots, X_n$, which are smaller than or equal to X_i.

Statistics of the form

$$S_n = \sum_{i=1}^{n} c_{ni} a_n(R_{ni}) \tag{2}$$

are called simple linear rank statistics. Here $(c_{n1}, \ldots, c_{nn})$ are given 'regression constants' and $(a(1), \ldots, a(n))$ denote the 'scores'. These scores are usually assumed to be generated by a function $\varphi(t),\ 0 < t < 1$; e.g. by simply setting

$$a_n(i) = \varphi\left(\frac{i}{n+1}\right), \ 1 \leq i \leq n \tag{3}$$

Another well-known choice is to take $a_n(i) = E\varphi(U_{i:n}),\ 1 \leq i \leq n$, where $U_{i:n}$ denotes the ith order statistic of a sample of size n from the uniform distribution on (0,1).

Simple linear rank statistics play a very important role in the nonparametric theory of rank tests. These statistics are useful in a variety of nonparametric testing problems, e.g. the two-sample problem (take c_{ni}=1 or 0), testing for randomness against trend, etc. In Hájek and Šidák [6] a wealth of asymptotic results on teststatistics of type S_n under the null-hypothesis $F_1 = \ldots = F_n$ and contiguous (local) alternatives can be found. In the important paper Hájek(1968) the asymptotic normality of S_n under *general* alternatives $(F_1, \ldots, F_n)$ was investigated.

Let us now briefly review the interesting methodology of the latter paper. Hájek's idea is first to approximate S_n by a suitable sum of independent random variables (r.v.) $L_n = \sum_{i=1}^n \ell_i(X_i)$, to which the classical central limit theorem can be applied. The functions ℓ_i are completely arbitrary, but must satisfy $E\ell_i^2(X_i) < \infty$. Here and elsewhere E denotes the expectation operator, while $E(\ |X)$ refers to the conditional expectation given X. The best choice for L_n is to take the 'projection' of S_n on the subspace $\mathcal{L}$ of all statistics L_n in the Hilbert space of all square integrable statistics based on $X_1, \ldots, X_n$. The resulting 'projection' $\hat{S}_n$ is easily found:

$$\hat{S}_n = \sum_{i=1}^{n} E(S_n|X_i) - (n-1)ES_n \tag{4}$$

where $ES_n = E\hat{S}_n$. In addition, one has the very useful relation:

$$E(S_n - \hat{S}_n)^2 = \sigma^2(S_n) - \sigma^2(\hat{S}_n) \tag{5}$$

whereas $E(S_n - L_n)^2 = E(S_n - \hat{S}_n)^2 + E(\hat{S}_n - L_n)^2$, so that indeed $\hat{S}_n$ is 'best possible' in a 'mean square' sense. (Here and elsewhere $\sigma^2(Z)$ denotes the variance of a r.v. Z.) With the aid of this simple projection argument Hájek was able to prove the asymptotic normality of S_n for score generating functions φ with a bounded second derivative, provided degeneration of the variance $\sigma^2(S_n)$ is avoided. This is achieved by proving that the 'remainder'term $S_n - \hat{S}_n$ is of negligible order of magnitude.

The second important step in Hájek's proof is to remove the restrictive requirement that φ possesses a bounded second derivative. In fact Hájek showed that this assumption is completely superfluous and can be replaced by the much weaker assumption that φ can be expressed as $\varphi = \varphi_1 - \varphi_2$, where the φ_i's (i=1,2) are both nondecreasing, square integrable and absolutely continuous inside (0,1). To prove this Hájek established the following intriguing inequality:

$$\sigma^2(S_n) \leq\ 21 \max_{1\leq i\leq n} (c_{in} - \bar{c}_n)^2 \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2 \tag{6}$$

where $\bar{c}_n = n^{-1}\sum_{i=1}^n c_{in}$ and $\bar{a}_n = n^{-1}\sum_{i=1}^n a_n(i)$. The c_{in}'s are arbitrary constants and $a_n(1) \leq \ldots \leq a_n(n)$ are non-decreasing constants. No condition (except continuity) is imposed on the underlying df's $F_1, \ldots, F_n$. With the aid of this variance inequality Hájek showed in fact the following: Suppose asymptotic normality of S_n is proved for a certain class $\Phi_0 = \{\varphi_0\}$ of score-generating functions, then it also holds for the class $\Phi = \{\varphi\}$ consisting of functions φ possessing the following property: for every $\epsilon > 0$, there exists a $\varphi_0 \in \Phi_0$ and non-decreasing functions φ_1 and φ_2 such that $\varphi = \varphi_0 + \varphi_1 - \varphi_2$ and $\int_0^1 (\varphi_1^2 + \varphi_2^2)\, dt\ < \epsilon$. If we take for Φ the class of all score-generating functions φ on (0,1) which can be expressed as $\varphi = \varphi_1 - \varphi_2$, where the φ_i's (i=1,2) are both nondecreasing, square integrable and absolutely continuous inside (0,1), then it turns out that Φ_0 may be taken as the class of all polynomials. This choice of Φ_0 is related to the well-known fact that the set of all polynomials is a dense subset of the L_1-space of integrable functions. Since any polynomial $\varphi_0 \in \Phi_0$ possesses a bounded second derivative, Hájek's proof of the asymptotic normality of S_n is now complete. The result of HÁJEK [7] was extended by DUPAČ AND HÁJEK [2] to the case of discontinuous φ, while in HOEFFDING [11] it was shown that the centering constant ES_n employed by Hájek in his 'asymptotic normality' result, can be replaced by a simpler, more practical, constant μ, provided the square integrability condition on φ_1 and φ_2 is slightly strengthened.

Hájek's method of proof was subsequently succesfully applied by STIGLER [14, 15] to the problem of establishing the asymptotic normality of L-statistics (or 'linear combinations of order statistics'). In addition, I should perhaps also note at this point that Hájek's 'projection' idea is nothing but a very special case of the well-known Hoeffding decomposition for general statistics with finite second moment. This decomposition was in recent years employed by many researchers (including the present author) in problems connected with the rate of convergence to normality, Edgeworth expansions and jackknife/bootstrap resampling methods.

H. Witting spoke about some old and new results concerning rank tests for scale in the nonparametric two-sample problem obtained at Freiburg during the past 15 years. Two Ph.D theses (by W. SCHÄFER (1979) and H.U. BURGER (1991)) were written on this topic. Several dispersion orderings and their applications to nonparametric tests were discussed. A nice review of all this is given in WITTING [18]. C. Domański (Lodz) discussed problems concerning the approximation of critical values of the well-known Wilcoxon two-sample rank test for location in the case of discrete distributions; i.e. with probability one ties are present.

Marie Hušková (Prague) gave a talk on a rank statistics approach to generalized bootstrap resampling schemes. The interesting basic idea, due to MASON AND NEWTON [13], is as follows: let $X_1, \ldots, X_n$ be a random sample of size n from the distribution function F, let $\theta(F)$ be a parameter of interest and $\theta(\hat{F}_n)$ its estimator, where $\hat{F}_n$ denotes the empirical distribution function based on $X_1, \ldots, X_n$. Efron's by now classical bootstrap is to take bootstrap samples from the distribution $\hat{F}_n$; the corresponding bootstrapped empirical distribution function is then given by

$$\hat{F}_n^*(x) = n^{-1} \sum_{i=1}^{n} I\{X_i \leq x\} M_i \tag{7}$$

where $I\{A\}$ denotes the indicator of the set A and $(M_1, \ldots, M_n)$ has the multinomial distribution $M(n; \frac{1}{n}, \ldots, \frac{1}{n})$. MASON AND NEWTON [13] proposed to replace $(M_1, \ldots, M_n)$ by weights $(nw_{n1}, \ldots, nw_{nn})$, that are nonnegative *exchangeable* random variables, satisfying certain mild conditions. The conditional distribution (given $X_1, \ldots, X_n$) of $\sum_{i=1}^{n} I(X_i \leq x) w_{ni}$ is now the same as the (conditional) one of $\sum_{i=1}^{n} I(X_i \leq x) w_{nR_i}$ (The conditioning here is on both $X_1, \ldots, X_n$ and $w_{1n}, \ldots, w_{nn}$; the randomness comes only through the ranks $R_1, \ldots, R_n$), where $R_1, \ldots, R_n$ is a random permutation of $1, \ldots, n$. In this way the problem of the consistency of such generalized weighted bootstrap resampling schemes reduces to the study of the limit behaviour of a certain class of rank tests. Hájek's 1961 ([5]) theorem on the limit distributions of simple linear rank statistics turns out to be a very powerful tool here. Results for specific statistics $\theta(\hat{F}_n)$ such as the sample mean, U-statistics and sample quantiles were presented.

3. Efficiency problems in estimation theory

Two famous results of Hájek - the convolution theorem and the LAM (local asymptotic minimax) theorem - obtained in the early seventies, became milestones in this area. This topic plays an important role in present day statistics, (see, e.g., Van der Vaart [16], LeCam [12]). While Hájek considers estimation of a parameter in 'smooth' parametric models $\{P_\theta, \theta \in \Theta\}$ (the parameter θ takes values in Θ, an open subset of $\mathcal{R}^k$), recent work in this area concentrates on nonparametric and semiparametric models.

Rudy Beran (Berkeley) gave an invited talk on the Hájek-LeCam convolution theorem, the LAM theorem and its connection with superefficiency, Stein estimation, and the bootstrap. Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n from P_θ and let θ be a vector-valued parameter, which we want to estimate. Let $T_n = T_n(X_1, \ldots, X_n)$ denote an estimator of θ, and let $H_n(\theta)$ denote the distribution of $\sqrt{n}(T_n - \theta)$. The sequence $\{T_n\}$ is called Hájek *regular* at θ_0 if $H_n(\theta_n) \Rightarrow H(\theta_0)$, as $n \to \infty$, whenever $\theta_n = \theta_0 + n^{-\frac{1}{2}}h$, for any fixed $h \in \Re^k$. Here $\Rightarrow$ denotes weak convergence. In Hájek [8] it is shown that if the model satisfies the LAN (Local Asymptotic Normality) property at θ_0 and $\{T_n\}$ is Hájek *regular* at θ_0, then $H(\theta_0)$ is the convolution of an optimal normal distribution and some other distribution (depending on the sequence $\{T_n\}$ at hand). In other words: asymptotically, for any regular estimator $T_n, \sqrt{n}(T_n - \theta)$ can be viewed as the sum of a normally distributed random variable - with variance equal to the inverse of the Fisher information number (Cramér-Rao bound) - and an independent 'noise' variable. As a consequence, an estimator sequence is called asymptotically efficient (i.e. 'best possible' within the class of all regular estimators) when it has a limiting normal distribution, with variance equal to the Cramér-Rao bound.

Regularity excludes superefficient estimators, such as the famous Hodges example – as it was presumably intended to do – and the well-known Stein shrinkage estimators. At this point one should recall that, when estimating the mean of a k-variate normal distribution, the Stein shrinkage estimator is admissible for quadratic loss functions, while the classical sample mean is inadmissible, if the dimension $k \geq 3$. So, perhaps lack of regularity is not always a bad thing.

In Hájek [9] the celebrated LAM-theorem is established. It asserts that if the model satisfies the LAN property one cannot estimate θ any better than in the normal limit problem. The lowerbound is in terms of risks (instead of distribution functions, as in the convolution theorem) and is valid for *any* estimator sequence $\{T_n\}$. In addition Hájek showed that, when k=1, any estimator sequence $\{T_n\}$ satisfying LAM – i.e. attaining the lower bound in terms of risks - is necessarily regular. When $k \geq 3$, however,this is no longer true any more: The Stein shrinkage estimators also achieve the lower bound; i.e. the improvement over the sample mean is not captured by LAM alone.

A nice connection between Hájek regularity and bootstrap resampling was also discussed by Beran. The *bootstrap estimator* of $H_n(\theta)$ is simply given by $H_n(\hat{\theta}_n)$, where $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ estimates θ consistently. Beran showed that if you have LAN at θ_0 and $H_n(\theta_n) \Rightarrow H(\theta_0)$, whenever $\sqrt{n}(\theta_n - \theta_0) \rightarrow h \in \Re^k$ and the sequence $\{\sqrt{n}(\hat{\theta}_n - \theta_0)\}$ is tight at θ_0, then $H_n(\hat{\theta}_n) \Rightarrow H(\theta_0)$ in probability under θ_0; i.e. the bootstrap estimator $H_n(\hat{\theta}_n)$ is asymptotically consistent. The assumption on $\{H_n(\theta_n)\}$ implies regularity and hence $H(\theta_0)$ has convolution structure. So, there appears to be a close link between bootstrap convergence and an appropriate 'stochastic regularity' condition. Beran also proved that if you have LAN at θ_0, $H_n(\hat{\theta}_n) \Rightarrow H(\theta_0)$ in probability under θ_0, and $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow Y$ under θ_0, where the random variable Y has full support in $\mathcal{R}^k$, then $H(\theta_0)$ has convolution structure. Extensions to nonparametric models (infinite dimensional parameter case) were also discussed. We note in passing that section 3.2 of the article by M.C. van Pul in this volume contains a somewhat similar consistency result for parametric bootstrapping in the context of software reliability.

Chris Klaassen (Amsterdam) discussed a result, which is, in a way, in between the LAM-theorem and the convolution theorem: no assumptions on the estimator sequence $\{T_n\}$ but still an assertion (though one which is a bit weaker than a convolution statement) in terms of distributions, like in the Hájek convolution theorem. I. Vincze (Budapest) spoke about some questions in connection with a non-regular Cramér-Rao type inequality.

4. Sampling from a finite population

The heading of this section is the title of another (posthumously published) book of Hájek ([10]).

P.K. Sen (Chapel Hill) gave an invited lecture on Hájek's asymptotic results for finite population sampling, such as rejective sampling and sampling with varying probabilities, and their extensions developed later on.

Zuzana Prášková (Prague) showed how a conjecture of Hájek (Conjecture 14.1 in [10]), concerning the relation between rejective sampling and conditional Poisson sampling, for the case when the population is divided into several strata, can be resolved in the affirmative. More precisely: Let S be a population of N units,$s \subset S$ a sample and P a probability distribution defined on the set of all subsets of S. Let $I_i(s)$ denote the inclusion indicator of unit i and let $\pi_i = EI_i(s)$ be the probability of inclusion of the unit i in the sample s. Poisson sampling with parameters $p_1, \ldots, p_N$ is the probability sampling scheme in which the inclusion indicators are independent zero-one random variables, with probabilities of inclusion $\pi_i = p_i$, for $i = 1, \ldots, N$. Rejective sampling of size n can be viewed as conditional Poisson sampling, under the condition that the sample size is fixed and equal to n. An asymptotic approximation of the inclusion probabilities π_i of the rejective sampling scheme in terms of $p_1, \dots, p_N$ was given by HÁJEK([10]). The same problem for the case that the population S is divided into strata $S_1, \dots, S_m$ was considered by Prášková. Let $K_h = |s \cap S_h|,\ h = 1, \dots, m$ denote the sample sizes in the strata $S_h,\ h = 1, \dots, m$. It can easily be checked that the including probabilities π_i can be written as

$$\pi_i = p_i \frac{P(K_1 = n_1, \dots, K_m = n_m \mid I_i = 1)}{P(K_1 = n_1, \dots, K_m = n_m)} \ ,\ i = 1, \dots, N$$

where $n_1, \dots, n_m$ are the fixed strata sample sizes. Prášková now establishes a multivariate Edgeworth expansion for the probabilities $P(K_1 = n_1, \dots, K_m = n_m)$ and combined with a similar result for $P(K_1 = n_1, \dots, K_m = n_m \mid I_i = 1)$ obtains her result.

The present author gave a talk on 'wild bootstrapping in finite populations'. This is work in progress at CWI joint with Marten Wegkamp (Leiden). The basic probabilistic tool we employ in our analysis is the celebrated Erdös-Rényi central limit theorem for samples drawn without replacement from a finite population. HÁJEK [4] showed that the Lindeberg type condition needed for asymptotic normality here is not only sufficient but necessary as well.

5. MISCELLANEOUS

J. Anděl (Prague) discussed a Bayesian approach to the periodogram and J. Dupačová (Prague) gave a talk entitled: 'Hájek and optimization'. J. Štěpán (Prague) spoke about a problem formulated by Hájek in 1969: What are geometrical and measure theoretical properties to characterize the Borel sets D in the unit square that support at most one probability measure P with given marginals P_1 and P_2? What is the relation of these sets to extremal measures in the compact convex set of all probability distributions with marginals P_1, P_2? The other talks dealt with a problem in numerical analysis, namely the computation of stationary probability vectors of large scale stochastic matrices (P. Mayer (Prague); joint work with I. Marek (Prague)) and confidence sets of fixed form and size with predetermined level of confidence (S. Holm (Göteborg)). In his lecture Sture Holm discussed an exact three stage method for obtaining confidence sets of a given general form and size in the ANOVA model with normally distributed observations. This is related to the classical two-stage Stein method for obtaining a confidence interval of fixed length and given confidence level for the mean of a normal distribution with unknown variance.

ACKNOWLEDGEMENTS

I am grateful to the participants of the Minisymposium for sending me summaries of their talks.

REFERENCES

1. VÁCLAV DUPAČ (1975), Jaroslav Hájek, 1926–1974, *The Annals of Statis-*

tics, **3**, 1031–1037.
2. V. DUPAČ AND J. HÁJEK (1969), Asymptotic normality of simple linear rank statistics under alternatives II, *The Annals of Mathematical Statistics*, **40**, 1992–2017.
3. C. GUTTENBRUNNER AND J. JUREČKOVÁ (1992), Regression rank scores and regression quantiles, *The Annals of Statistics*, **20**, 305–330.
4. J. HÁJEK (1960), Limiting distributions in simple random sampling from a finite population, *Periodica Mathematica*, **5**, 361–374.
5. J. HÁJEK (1961), Some extensions of the Wald-Wolfowitz- Noether theorem, *The Annals of Mathematical Statistics*, **32**, 506–523.
6. J. HÁJEK AND Z. ŠIDÁK (1967), *Theory of Rank Tests.* Academia, Prague and Academic Press, New York. (Russian translation: IZD. NANKA, Moscow 1971).
7. J. HÁJEK (1968), Asymptotic normality of simple linear rank statistics under alternatives, *The Annals of Mathematical Statistics*, **39**, 325–346.
8. J. HÁJEK (1970), A characterization of limiting distributions of regular estimates, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **14**, 323–330.
9. J. HÁJEK (1972), Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob. 1*, 175–194. Univ. of California Press.
10. J. HÁJEK (1981), *Sampling from a finite population*, Marcel Dekker, Inc, New York (edited by VÁCLAV DUPAČ).
11. W. HOEFFDING (1973), On the centering of a simple linear rank statistic, The Annals of Statistics, **1**, 54–66.
12. L. LECAM (1994), An infinite dimensional convolution theorem, in: *Stat. Dec. Theor. and Rel. Topics V* (eds. S.S. GUPTA, J.O. BERGER), 401–411, Springer.
13. D.M. MASON AND M.A. NEWTON (1992), A rank statistic approach to the consistency of a general bootstrap, The Annals of Statistics, **20**, 1611–1624.
14. S.M. STIGLER (1969), Linear functions of order statistics, The Annals of Mathematical Statistics, **40**, 770–788.
15. S.M. STIGLER (1974), Linear functions of order statistics with smooth weight functions, The Annals of Statistics, **2**, 676–693.
16. A. VAN DER VAART (1991), An asymptotic representation theorem, Internat. Statist. Review **58**, 97–121.
17. H. WITTING AND G. NÖLLE (1970), *Angewandte Mathematische Statistik, Optimale finite und asymptotische Verfahren*, B.G. Teubner Stuttgart.
18. H. WITTING (1994), Rank tests for scale: Hájek's influence and recent developments, to appear in *Kybernetika*.