STATISTICS 592B
SEMIPARAMETRIC MODELS

Winter Quarter 1998 (3 Credits)

TTh 10:30 am – 12:00 noon, Padelford C301

INSTRUCTOR:
Jon A. Wellner
B320 Padelford
543-6207; jaw@biostat.washington.edu
Office hours: 1:30 - 3:30 MWF (or by appointment)

Web Address:
http://www.stat.washington.edu/jaw/COURSES/592B.98/98.semip.html

TEXTS:

- *Efficient and Adaptive Estimation for Semiparametric Models*, by P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, Johns Hopkins University Press, Baltimore (1993). [Soon (2/98?) to be reprinted in paperbook form by Springer-Verlag.]
- *Information Bounds and Nonparametric Maximum Likelihood Estimation*, by P. Groeneboom and J. A. Wellner, Birkhauser, (1992).

PREREQUISITES: STAT 583 (or permission of instructor)

COURSE FORMAT AND REQUIREMENTS: The instructor will lecture about 80% of the course. The remaining time will be devoted to the discussions of recent research in journal literature on topics selected by students. The initial material covered in the first third to half of the lectures will be based on the two texts listed above, but much of the remaining material covered in the course will be based on recent journal literature or other books. Students will be expected to read appropriate book chapters and key articles related to their chosen topic. There will be a limited number of homework problems, and no written examinations.

HOMEWORK: Minimal. Handed out on Thursdays; due following Thursday.

PROJECTS: Paper, about 10 pages; plus a half-hour talk toward the end of the quarter. See the list of potential topics for some suggestions, or – better yet – base your paper and talk on a model of interest to you! I'm still adding to the topics list on the course web-page; check there again later this week.

GRADING:

- Homework: 35%
- Project: 65%

STATISTICS 592B
SEMIPARAMETRIC MODELS

TENTATIVE COURSE OUTLINE:

1. Information Bounds
   - Background theory: contiguity
   - Parametric Models
   - Nonparametric Models and Semiparametric Models
   - Differentiability of Implicitly Defined Functions
2. Construction of Estimates: General Theory
   - Bickel-Klaassen-Schick theorem
   - Examples and Discussion
   - Ibragimov and Has'minskii theorems
3. M- and Z-estimation Results and Techniques (Estimating Equations)
   - Finite-dimensional Z-estimators: Le Cam, Pollard, BKRW
   - Infinite-dimensional Z-estimators: BKRW, Van der Vaart
   - M-estimation and rates of convergence: Birgé and Massart, Wong and Shen, VdV and Wellner
4. Maximum Likelihood and Pseudo-Likelihood Estimation in Smooth Semiparametric Models: Empirical process methods
   - Huang's theorem: Examples
   - Theorems of Newey and Hu
   - Van der Vaart's theorem for Mixture Models

STATISTICS 592B
SEMIPARAMETRIC MODELS

OTHER BOOKS and BASIC REFERENCES

BOOKS:

(BKRW) Bickel, P., Klaassen, C.A.J., Ritov, Y., and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

(GW) Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkauser, Basel.

(IH) Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

(LC) Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

(LCY) Le Cam, L. and Yang, G. L. (1990). *Asymptotics in Statistics, Some Basic Concepts.* Springer-Verlag, New York.

(M) Millar, P. W. (1983). The Minimax Principle in Asymptotic Statistical Theory. In *Ecole d'Eté de Probabilités de Saint-Flour XI-1981. Lecture Notes in Mathematics* **676**, 76 - 262, Springer-Verlag, New York.

(PfI) Pfanzagl, J. (with the assistance of W. Wefelmeyer) (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statistics* **13**, Springer-Verlag, New York.

(PfII) Pfanzagl, J. (1990). *Estimation in Semiparametric Models: Some Recent Developments. Lecture Notes in Statistics* **63**, Springer-Verlag, New York.

(VdV) Van der Vaart, A. W. (1988). *Statistical estimation in large parameter spaces.* CWI, Math. Centrum, Centrum voor Wiskunde en Informatica, Amsterdam.

(VW) Van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

STATISTICS 592B
SEMIPARAMETRIC MODELS

Adaptive Estimation, the Hájek - Le Cam Convolution Theorem, and (Local) Asymptotic Minimax Theorems

A history and annotated bibliography

I. Adaptive Estimation.

1956 Stein, C. Efficient nonparametric estimation and testing. *Proc Third Berkeley Symp. Math. Statist. Prob.* **1**, 187-195.
Stein points out the possibility of "adaptive" estimation in estimating the center (i.e. location) of a symmetric density and in estimation of shift in the two-sample problem.

1967 Hájek, J. and Zidak, Z. *Theory of Rank Tests.* Academic Press, New York.
In Section VII.1.6 adaptive rank tests are constructed which are asymptotically efficient in the two-sample shift model.

1970 van Eeden, C. Efficiency-robust estimation of location. *Ann. Math. Statist.* **41**, 172 - 181.
The first construction of adaptive (efficient) estimators in the one-sample symmetry problem.

1974 Beran, R. Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2**, 63 - 74.
Beran constructs adaptive estimators in the one-sample symmetry problem based on ranks; adaptive "R-estimators".

1975 Sacks, J. An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.* **3**, 285 - 298.
Sacks constructs adaptive estimators in the one-sample symmetry problem based on linear combinations of order statistics; adaptive "L-estimators".

1975 Stone, C. J. Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3**, 267 - 284.
Stone constructs adaptive estimators in the one-sample symmetry problem based on solving the "estimated score equation"; i.e. adaptive "M-estimators".

1982 Bickel, P. J. On adaptive estimation. *Ann. Statist.* **10**, 647 - 671.
Bickel reviews the state of adaptive estimation as of 1980 and extends adaptive estimation to several group models.

II. The convolution theorem.

1970 J. Hájek. A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie* **14**, 323 - 330.
First statement of the result for the finite - dimensional or "parametric" $\theta \in R^d$ case; a "Bayesian proof" is given under a local asymptotic normality (LAN) assumption (not restricted to iid).

1970 Inagaki, N. On the limiting distribution of a sequence of regular estimates. *Ann. Inst. Statist. Math.* **22**, 1 - 13.
A theorem almost identical to Hájek's is given.

1972 George Roussas. *Contiguity of Probability Measures: some applications to statistics.* Cambridge University Press.
A proof due to Bickel based on complex analysis is given: an analytic function which is constant on the real axis is constant. The convolution theorem, in terms of transforms, is then read off the imaginary axis. (Roussas p. 136).

1972 Hájek, J. Local asymptotic minimax and admissibility in estimation. *Proceedings of the Sixth Berkeley Symposium*, **1**, 175 - 194.
Contains a very nice exposition of the convolution theorem, the rationale for large - sample statistical theory, and the related work by Le Cam as well as a nice formulation of several local asymptotic minimax theorems in a spirit similar to that of the convolution theorem.

1972 L. Le Cam. Limits of experiments. *Proceedings of the 6th Berkeley Symposium*, **1**, 245 - 261.
A convolution theorem for the parametric case more general than Hájek's is given. The proof is based on the Markov - Kakutani fixed - point theorem and the Tulcea lifting theorem.
Correction: L. Le Cam. On a theorem of J. Hájek. In *Contributions to Statistics; Jaroslav Hájek Memorial Volume.* J. Jureckova, ed., Reidel, 119 - 135.

1977 Beran, R. Robust location estimates. *Ann. Statist.* **5**, 431 - 444.
Beran, R. Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5**, 445 - 463.
Beran, R. Estimating a distribution function. *Ann. Statist.* **5**, 400 - 404.
In these three papers convolution theorems are given for the types of estimates indicated in the titles. Fréchet differentiability of functionals wrt the Hellinger metric plays a key role, and the proofs are of the complex variable type. The third paper gives what is apparently the first "infinite - dimensional" convolution theorem.

1981 Ibragimov, I. A. and Hasminskii, R. Z. *Statistical Estimation: Asymptotic Theory.* Springer Verlag (original in Russian, 1979, Nauka).
Theorem 9.1 p. 154 is Hájek's convolution theorem. Bickel's proof is given. Relations among various definitions of efficiency are discussed.

1982 Fabian, V. and Hannan, J. On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrschenlichkeitstheorie v. Geb.* **59**, 459 - 478.
They clarify Hájek's results with regard to attainability of the bounds, and show that some uniformity is needed in the LAN condition in order for the bounds of the convolution and local asymptotic minimax theorems to be attainable.

1982 Pfanzagl, J. (with the assistance of W. Wefelmeyer). Contributions to a General Asymptotic Statistical Theory. *Lecture Notes in Statistics* **13**, Springer - Verlag, New York.
Pfanzagl and Wefelmeyer pursue the nonparametric point of view of Beran (1977) and the Russian group including Koshevnik and Levit (1976) and Levit (1975) by presenting convolution theorems for differentiable functionals, but allowing the possibility that the model is *not* fully nonparametric.

1983 Millar, P. W. The minimax principle in asymptotic statistical theory. Ecole d'Eté de Probabilités de St. Flour XI - 1981. *Lecture Notes in Math.* **876**, 76 - 262. Springer-Verlag, New York.
A version of Le Cam's infinite - dimensional convolution theorem is proved, and many examples given.

1983 Begun, J., Hall, W.J., Huang, W. M., and Wellner, J. A. Information and asymptotic efficiency in parametric - nonparametric models. *Ann. Statist.* **11**, 432 - 452.
The convolution theorem is extended to "semiparametric models"; the proofs are based on the characteristic function methods of Bickel and Beran.

1985 Millar, P. W. Nonparametric applications of an infinite dimensional convolution theorem. *Z. Wahrscheinlichkeitstheorie* **68**, 545 - 556.
A convolution theorem of Le Cam's type is given for experiments indexed by a Hilbert space. The proof uses Le Cam's theorem for the finite - dimensional subspace and extends the resulting measure from cylinder sets. Many applications to nonparametric estimation are given.

III. Asymptotic Minimax Theorems

1953 Le Cam, L. On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Statist.* **1**, 277 - 330.
Le Cam proves (theorem 14) that local asymptotic minimax implies local asymptotic admissibility, and that superefficiency excludes the local asymptotic minimax property.

1956 Chernoff, H. Large sample theory: Parametric case. *Ann. Math. Statist.* **27**, 1 - 22.
Chernoff discusses the local asymptotic minimax formulation as providing one way of rescuing Fisher's programs, and attributes the idea to C. Stein and H. Rubin.

1956 Dvoretzky, A., Kiefer, J., and Wolfowitz, J. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642 - 669.
Asymptotic minimaxity of the empirical df in the case of $\mathcal{X} = R$ is proved in a very strong sense.

1959 Kiefer, J. and Wolfowitz, J. Asymptotic minimax character of the sample distribution for vector chance variables. *Ann. Math. Statist.* **30**, 463 - 489.
The 1956 result for $\mathcal{X} = R$ is extended to $\mathcal{X} = R^d$ with $d \geq 2$; dificulties concerning uniformity of convergence of the empirical df are met head-on.

1976 Kiefer, J. and Wolfowitz, J. Asymptotic minimax estimation of concave and convex distributions. *Z. Wahrscheinlichkeitstheorie* **34**, 73 - 85.
It is shown that the empirical df remains asymptotically minimax for estimation of a convex or concave df; the distance between the greatest concave majorant of the empirical and the empirical df itself is analyzed closely.

1979 Le Cam, L. On a theorem of J. Hajek. In *Contributions to Statistics; Jaroslav Hajek Memorial Volume*, J. Jureckova, ed., Reidel, 119 - 135.

1979 Millar, P. W. Asymptotic minimax theorems for the sample distribution function. *Z. Wahrscheinlichkeitstheorie* **48**, 233 - 252.
Continuing the theme of the 1976 Kiefer and Wolfowitz paper, the sample distribution function is shown to be asymptotically minimax for families of distribution functions satisfying monotonicity or convexity restrictions, extending the classical results of Dvoretzky, Kiefer, and Wolfowitz (1956).

1983 Millar, P. W. The minimax principle in asymptotic statistical theory. Ecole d'Eté de Probabilités de St. Flour XI - 1981. *Lecture Notes in Math.* **976**, Springer - Verlag, 76 - 265.