Research Statement

Harrison H. Zhou
Cornell University

My research interests and contributions are in the areas of model selection, asymptotic decision theory, nonparametric function estimation and machine learning.

Model Selection

(**1**). *Minimax Estimation with Thresholding and its Application to Wavelet Analysis* (*with J. T.Gene Hwang*, 2003, *revised for Ann. Stat.*)

Many statistical problems involve selecting a model (a reduced model from the full model) and use it to do estimation. Is it possible to do so and still come up with an estimator always better than the naive estimator without model selection? The James-Stein estimator allows us to do so. However, the James-Stein estimator considers only one reduced model, the origin (cf. James and Stein, 1961). What should be more desirable is to select a data chosen reduced model (of an arbitrary dimension) and then do estimation. In Zhou and Hwang (2003), we construct such estimators. We apply the estimators to the wavelet regression. In finite sample settings, these estimators are minimax and perform the best among a class of well-known estimators (cf. Donoho and Johnstone, 1994, 1995, Cai, 1999, etc.) which do model selection and estimation at the same time. Some of our estimators are also shown to be asymptotically optimal.

(**2**). *Minimax Variable Selection in Linear Regression (with J. T. Gene Hwang, forthcoming)*

We extend the minimax estimators in Zhou and Hwang (2003) to normal linear models under mean squared error loss or predictive mean squared error loss. The ordinary least square (OLS) is the naive estimator based on the full model. Our estimator truncates the components of OLS which are small (i.e., give zero as the estimate when the OLS estimate has a small magnitude) and preserves (or precisely shrinks) the others. Simulation studies show the estimator improves substantially upon the naive estimator (OLS) when the reduced model is correct. The novelty here as compared to other estimators (even to Zhou and Hwang (2003)) is that our estimator is designed to do truncation even when the covariance is nondiagonal.

Following up on this project (joint with J.T. Gene Hwang): we will apply this minimax estimation approach to an ANOVA model in microarray data analysis (cf. Kerr, Martin, and Churchill, 2000).

Asymptotic Decision Theory

The following projects are supported by the National Science Foundation with Grant DMS-0306497: *Asymptotic equivalence of statistical experiments*, 2003-2008.

Asymptotic equivalence has been shown by Le Cam (1986) to provide the adequate framework for decision theoretic limit theorems. For finite dimensional parameter spaces, sequences of rescaled (localized) experiments around a given parameter converge to a Gaussian limit (local asymptotic normality, LAN). As a consequence, optimal solutions for large classes of statistical decision problems (estimation, testing) can be derived from the limit experiment. The LAN-theory has become a centerpiece and a standard tool in asymptotic statistics (cf. van der Vaart, 1998, Shiryaev and Spokoiny, 2000). An extension to infinite dimensional parameter spaces and to a global approximation (without localization), in the context of i.i.d. experiments has been given in Nussbaum (1996). Further recent developments (Carter, 2001, 2002, Brown, Cai, Low and Zhang, 2002, Wang, 2002, Brown, Carter, Low and Zhang, 2003, etc.) justify a claim that asymptotic equivalence theory is emerging as a recognizable research area in statistics. The following are two forthcoming works.

(**3**). *Asymptotic Equivalence of Spectral Density and Gaussian White Noise (with G. Golubev, M. Nussbaum, forthcoming)*

Dzhaparidze (1986) presents a comprehensive treatment of inference for spectral densities of Gaussian stationary time series, based in part on the LAN-approach. The validity of the local approximation by Gaussian experiments in parametric cases suggests a global white noise approximation for nonparametric sets of spectral densities. We establish asymptotic equivalence in the sense of Le Cam's deficiency distance to the problem of signal estimation in Gaussian white noise where the signal is log-spectral density. The first step of the proof is the reduction of the stationary series to independent Gaussians with unkown, smoothly varying variances. That nonparametric Gaussian scale model can be reduced to a Gaussian location model via a multiresolution scheme involving Beta distributions, in the spirit of the Hungrian construction for empirical processes. The asymptotic equivalence result is established over a Besov (Sobolev) type space.

(**4**). *Poissonization of I.I.D. Experiments (with M. G. Low, forthcoming)*

Here the aim is to show that the experiments given by observations

$$\begin{aligned}&y_1, y_2, \ldots, y_n \ i.i.d. \text{ with law } \mathcal{P}\text{on } \Omega\\&X_n(\cdot), \text{a Poisson process on } \Omega \text{ with intensity measure } n\mathcal{P}\end{aligned}$$

are asymptotically equivalent over a parameter space (class of laws) $P \in \mathcal{P}$. Le Cam (1974) proved that deficiency distance converges to 0 with rate $n^{-1/4}$ for parametric sets $\mathcal{P}$; Mammen (1986) improved this rate to $n^{-1/2}$. In Le Cam (1986), p. 508, conditions are given for general nonparametric sets $\mathcal{P}$ in terms of Hellinger metric entropy; these conditions specialize to a bound for smoothness $\alpha > 1/2$ (cf. also Le Cam and Yang (2000), p. 73). This Poissonization result for nonparametric i.i.d. experiments was used as a technical tool for the Gaussian approximation in Nussbaum (1996). Based on the rationale that a Poisson approximation should be valid under substantially weaker conditions than a Gaussian one, the question can be asked whether Le Cam's smoothness bound $\alpha > 1/2$ is sharp. For densities on $[0, 1]$ we give a sharp Besov smoothness condition $\alpha p > 1/2$ for poissonization. The following quotation, referring to Gaussian and Poisson approximations, offers an interesting perspective on our program: "Poisson experiments are less tractable and less studied. They will loom large in the new century" (Le Cam and Yang (2000), Preface to the Second Edition). The program of poissonization can be seen as a first step towards an infinitely divisible approximation.

Following up on this project – Infinitely Divisible Approximations for Nonparametric i.i.d. Experiments (joint with Nussbaum): The asymptotic equivalence of density estimation and Gaussian white noise has been established in Nussbaum (1996), where for a nonparametric class of densities a common support [0,1] is assumed. The question we pose here is: Can the condition "bounded away from zero" be weakened or removed? It can be seen that if an additional location parameter is introduced, then the Gaussian white noise approximation fails. This paper shows that the correct approximation in this case is an infinitely divisible experiment that is a Gaussian/Poisson mixture. In analogy to results for endpoint estimation, the tail rates of the densities are crucial for the shape of the approximation. A theory of infinitely divisible experiments has been developed in the monograph of Janssen, Milbrodt and Strasser (1985), with a view to parametric models and local limits. Our focus here is on nonparametric i.i.d. models and global asymptotic equivalence (cf. Nussbaum and Zhou, Talk at Purdue, 2003)

*Future Research***:** We will study explicit Markov kernels to establish asymptotic equivalence theory for generalized linear models and location type regression model . The LAN type results have been established for many models, including long memory process and hidden Markov models (cf. Hallin et al, 1999, Bickel and Ritov, 1996, Golubev and Khasminski, 1998). That suggests a global asymptotic equivalence theory for those stochastic models. The applications of asymptotic equivalence theory are very promising. They include looking for optimal constants and convergence rates in minimax estimation (cf. Korostelev and Nussbaum, 1999) , and precise Kolmogrov ε-entropy and Kolmogrov complexity in information theory (cf. Donoho, 2003, Le Cam Lecture, Donoho, 2002, Yang and Barron, 1999), etc.

Function Estimation

(**5**). *SURE Approach to Block James-Stein Thresholding in Wavelet Regression (with T. T. Cai, forthcoming).*

The wavelet methodology has demonstrated considerable success in terms of spatial adaptivity and asymptotic optimality. In particular, block thresholding rules have been shown to possess impressive properties. The estimators make simultaneous decisions to retain or to discard all the coefficients within a block and increase estimation accuracy by utilizing information about neighboring coefficients. The idea of block thresholding can be traced back to Efromovich (1985) in orthogonal series estimators. In the context of nonparametric regression local block thresholding has been studied in, for example, in Hall, Kerkyacharian, and Picard (1998), Cai (1999), and Efromovich (2002). In this joint paper with Cai we propose Sure+BlockJS approach, which chooses block sizes and thresholding levels adaptively, i.e., uses data to determine the block sizes and theresholding levels to be used. It can be proved that the IMSE of Sure+BlockJS is asymptotically better than methods proposed in Donoho and Johnstone (1994, 1995), Cai (1999), etc.

Future research – exact asymptotic adaptation for wavelet estimation: Exact adaptation in the minimax sense was introduced by Pinsker(1980). In this direction one can cite papers by Efromovich and Pinsker(1984), Nussbaum (1985), Golubev and Nussbaum (1992), Korostelev (1993), Donoho (1994), Beran (1996), Lepskii and Spokoiny (1997), Tsybakov (1997), etc. Nearly optimal and optimal convergence rates have been achieved for Besov class $B^{\alpha}_{p,q}$ when $p < 2$ (cf. Donoho and Johnstone,1994, 1995, 1998), but the exact adaptive minimax estimators are still unknown.

(**6**). *A Root-unroot Transform and Wavelet Block Thresholding Approach to Adaptive Density Estimation (with L. D. Brown, T. T. Cai, R. Zhang, L. H. Zhao among others, forthcoming).*

This paper describes an algorithm for nonparametric density estimation using a root-unroot paradigm. The paradigm involves several easily implemented steps as follows: suitably bin the data; calculate the square root of the normalized binned data; apply wavelet block thresholding approach. Then "unroot" in a suitable fashion. The binning step involves only an insignificant loss of information. It can be proved that this procedure can achieve the optimal minimax convergence rate adaptively over a broad range of Besov spaces using a quantile coupling inequality (cf. Komlos, Major, Tusnady, 1975). The methodology is equally suitable for nonparametrically estimating the intensity of an inhomogeneous Poisson process.

Following up on this project (joint with L. D. Brown and T. T. Cai)**:** Similar paradigms have been found for nonparametric generalized linear models. For a location type regression with heavy tail, we can also find a transformation to Gaussian regression.

Machine Learning

(**7**)**.** *Global Geometry of SVM Classifiers (with D. Zhou, B. Xiao, and R. Dai, 2002)*

We construct an alternative geometry framework for Support Vector Machine (SVM) classifiers. Within this framework, separating hyperplanes, dual descriptions and solutions of SVM classifiers are all constructed clearly by a purely geometric fashion. Now all kinds of SVM formulations and their dual descriptions including the arbitrary-norm cases are only different expressions of the underlying common geometric essentials. Compared with the optimization theory in SVM classifiers, we don't need redundant involved computations any more. Instead, every step in our theory is guided by elegant geometric intuitions. Our framework can make people understand SVM in a totally visual fashion. In addition, it is also helpful to expose the correlations between SVM and other learning algorithms.

References

[1] Beran, R. (1996). Stein estimation in high dimensions: a retrospective. In: Madan Puri Festschrift (E. Brunner and M. Denker, eds.) 91-110 . VSP Editors, Zeist (Netherlands)

[2] Bickel, P. J. and Ritov, Y. (1996). Inference in hidden Markov models I: Local asymptotic normality in the stationary case, Bernoulli **2** 199-228

[3] Brown, L. D., Cai, T. T., Low, M. G. and Zhang, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. Ann. Statist. **30** (3) 688-707

[4] Brown, L.D., Carter, A. V., Low M. G., and Zhang, C.-H. (2003) Asymptotic equivalence theory for density estimation, Poisson processes, and Gaussian with noise with drift. Ann. Statist. To appear.

[5] Brown, L. D. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. Ann. Statist. **24** 2384-2398

[6] Cai, T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. Ann. Statist. **27**, 898-924.

[7] Carter, A. V. (2001). Deficiency distance between multinomial and multivariate normal experiments under smoothness constraints on the parameter set. Technical Report, UC Santa Barbara. Available at www.pstat.ucsb.edu/faculty/carter/research.html.

[8] Carter, A. V. (2002). Deficiency distance between multinomial and multivariate normal experiments. Ann. Statist. **30** (3) 708-730

[9] Chentsov, N. N. (1982). Statistical Decision Rules and Optimal Inference. Translations of Mathematical Monographs, **53**. American Mathematical Society, Providence, R.I.

[10] Delattre, S. and Hoffmann, M. (2001). The Pinsker bound in mixed Gaussian white noise. Math. Methods Statist. **10** (2) 283-315

[11] Donoho, D. (1994). Asymptotic minimax risk (for sup-norm loss): solution via optimal recovery. Probab. Theor. Rel. Fields **99** 145-170.

[12] Donoho, D. L. (2002). The Kolmogorov sampler. Unpublished manuscript.

[13] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion) J. Roy. Statist. Soc. B **57** 301-369.

[14] Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation via wavelet shrinkage. Biometrika **81** 425-455.

[15] Donoho, D.L. & Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. J. Amer. Statist. Assoc. **90**, 1200-24.

[16] Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. Ann. Statist. **26** 879-921.

[17] Dzhaparidze, K. (1986). Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series. Springer-Verlag, New York-Berlin.

[18] Efromovich, S.Y. (1985). Nonparametric Estimation of a Density of Unknown Smoothness. Theor. Probab. Appl. **30**, 557-661.

[19] Efromovich, S.Y. (2002). On blockwise shrinkage estimation. Technical Report.

[20] Efromovich, S. Y. & Pinsker, M. S. (1984), Self learning algorithm of nonparametric filtration, (in Russian). Automat. i Telemeh. **11**, 58-65.

[21] Golubev, G. and Khasminskii, R. (1998). Asymptotically optimal filtering for a hidden Markov model. Math. Methods Statist. **7** 192-209.

[22] Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. Ann. Statist. **18**, 758-778.

[23] Grama, I.and Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. Probab. Theor. Related Fields **111** 493-533

[24] Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. Ann. Statist. **26** 922-942.

[25] Hallin, M., Taniguchi, M., Serroukh, A., and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance, Ann. Statist. **27** 2054-2080

[26] Janssen, A., Milbrodt, H. and Strasser, H. (1985). Infinitely Divisible Statistical Experiments. Lecture Notes in Statist. 27. Springer, New York.

[27] Kerr, M. K., Martin, M. A. and Churchill, G. A. (2000), Analysis of variance for gene expression microarray data, Journal of Computational Biology 7, 819-837,

[28] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample df. I Z. Wahrsch. verw. Gebiete **32** 111-131.

[29] Korostelev, A. P. (1993). An asymptotically minimax regression estimator in the uniform norm up to a constant. Theory Probab. Appl. **38**, no. 4, 775-782

[30] Korostelev, A. P. and Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation, Bernoulli **5** No. 6

[31] Le Cam, L. (1974). On the information contained in additional observations. Ann. Statist. **2** 630-649.

[32] Le Cam, L. (1986) Asymptotic Methods in Statistical Decision Theory. Springer-Verlag, New-York etc.

[33] Le Cam, L., and Yang, G.L. (2000). Asymptotics in Statistics. Some Basic Concepts, 2nd ed. Springer, New York.

[34] Lepskii, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. Ann. Statist. **25** 2512-2546

[35] Mammen, E. (1986). The statistical information contained in additional observations. Ann. Statist. **14** 665-678

[36] Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. Ann. Statist. **13** 984-997

[37] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise, Ann. Statist. **24** 2399-2430 (1996)

[38] Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise (in Russian). Problemy Peredachi Informatsii **16** (2) 52-68, English translation Problems Inform. Transmission (1980) 120-133

[39] Tsybakov, A. B. (1997). Asymptotically efficient signal estimation in L2 under general loss functions. Problems of Information Transmission **33** no. 1, 78-88.

[40] Van der Vaart, A. (1998). Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics, 3)

[41] Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and difiusions. Ann. Statist. **30** (3) 754-783

[42] Shiryaev, A. N. and Spokoiny, V. (2000). Statistical Experiments and Decisions: Asymptotic Theory. World Scientific, Singapore.

[43] Yang, Y and Barron, A (1999), Information-theoretic determination of minimax rates of convergence, Ann. Statist. **27** 1564-1599

[44] Zhou, D., Xiao, B., Zhou, H. H. and Dai, R. (2002). Global Geometry of SVM Classifiers. Unpublished manuscript.

[45] Zhou, H. H. and Hwang, J. T. G. (2003). Minimax estimation with thresholding and its application to wavelet analysis (revised for Ann. Stat.).