Andrew B. Nobel

Professor of Statistics and Biostatistics
Department of Statistics and Operations Research
University of North Carolina
Chapel Hill, NC 27599-3260

Fax: 919.962.1279
Telephone: 919.962.1352
E-mail: nobel@email.unc.edu

Education

Doctorate, Electrical Engineering, Stanford University, 1992

Masters, Electrical Engineering, Stanford University, 1988

Certificate of Study, Mathematics, Cambridge University, England, 1986

Bachelor of Science (with honors), Electrical Engineering, Cornell University, 1985

Academic Appointments

Assistant/Associate Professor, Department of Statistics, UNC Chapel Hill, 1994 - 2006

Professor, Department of Biostatistics, UNC Chapel Hill, 2007 -

Adjunct Professor, Department of Computer Science, UNC Chapel Hill, 2006 -

Faculty Fellow, Statistics and Applied Mathematical Sciences Institute (SAMSI), 2004-2005

Visiting Associate Professor, Dept. Statistics, University of Chicago, Fall Quarter, 2000

Visiting Assistant Professor, Department of Mathematics, Université de Paris-Sud, Orsay, May-June 2000

Postdoctoral Fellow, Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, 1992 - 1994

Summer Research Associate, AT&T Bell Labs, Murray Hill, NJ, 1990

HONORS

IMS Fellow, 2008

Lucent Distinguished Lecturer, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, March, 2002

National Science Foundation CAREER Grant, 1995

Beckman Institute Fellow, University of Illinois U-C, 1992-1995

IBM Doctoral Fellow, 1988

National Science Foundation Graduate Fellow, 1985

Churchill Scholar, Cambridge University, 1985

External Professional Activities

Electronic Journal of Statistics

Associate Editor, 2007-

Institute of Mathematical Statistics (IMS)

Program Secretary and Member of Executive Committee, 2003 - 2006

Member, Committee on Special Lectures, 2006-

IEEE Transactions on Information Theory

Associate Editor for Pattern Recognition, Statistical Learning and Inference, 2002 - 2004

Winston Churchill Scholarships

National Selection Committee, Chicago, IL, December, 2002.

Conferences:

Program Committee for the *Joint World Congress of the Bernoulli Society and Annual Meeting of the Institute for Mathematical Statistics*, National University of Singapore, 2008.

Program Committee for the *Third IMS-ISBA Joing Meeting*, Bormio, Italy, 2007.

Co-organizer of Classification Session at the *IEEE Information Theory Workshop on Detection, Estimation, Classification, and Imaging*, Santa Fe, NM, 1999.

Organizing Committee for the *IMS Special Topics Meeting, 50'th Anniversary Commemorative Conference*, Department of Statistics, Univerisity of North Carolina, Chapel Hill, 1996.

Statistics and Applied Mathematical Sciences Institute (SAMSI)

Local Development Committee, 2002-2003

Advisory Committee and Participant: Program on Computational Biology of Infectious Diseases

Reviewing:

Journals: Annals of Statistics, IEEE Transactions on Information Theory, Journal of Multivariate Analysis, Statistics and Probability Letters, Machine Learning, Applied and Computational Harmonic Analysis, Statistics, Statistica Neerlandica, Computational Statistics, Annals of Applied Probability, Annals of the Institute of Statistical Mathematics, Bernoulli, Bioinformatics, Dynamical Systems.

Books: Springer Verlag **Grant applications:** NSF and NSA

Select Current Funding

Carolina Environmental Bio-informatics Research Center
F.A. Wright, PI
EPA STAR RD-83272001
Grant period 2005-2010, total award $4,400,000

NCI Specialized Program of Research Excellence, Cores 1 and 2
H. Shelton Earp III, PI
NCI SPORE
Grant period 2006 - 2011

Select Past Funding

Analysis of High Dimensional Data Using Subspace Clustering
National Science Foundation, DMS-0406361
A.B. Nobel, PI, Wei Wang, Co-PI
Grant period 2004-2008

Discovering New Functional Relationships from Gene Expression Data with Structured Permutation Tests
Cystic Fibrosis Foundation Therapeutics, Inc., NOBEL04V0
A.B. Nobel, PI
Grant period 2004-2007

The Carolina Center for Exploratory Genetic Analysis
National Institutes of Health, P20 RR020751-01
Dan Reed, PI
Grant period 2004-2007

UNC Kenan Faculty Retention Fund, Arts and Sciences, 2004-2005

Generation and Validation of Synthetic Internet Traffic
National Science Foundation, ANI-0323648
K. Jeffay, PI, A.B. Nobel and F.D. Smith, Co-PI
Grant period 2003-2005, total award $470,000

Estimation from dynamical systems and individual sequences

A.B. Nobel, PI

National Science Foundation, DMS-9971964.

Grant period 1999-2003

Greedy Growing and its Applications

Faculty Early Career Development (CAREER) Award

A.B. Nobel, PI

National Science Foundation, DMS-9501926.

Grant period 1995-1998

Equipment Grant for Scientific Computing (SCREMS)

National Science Foundation, DMS-9506575.

Joint with Professors J. Fan, C. Ji, and S. Marron.

Grant period 1995-1996

PUBLICATIONS

A. Refereed Publications

A Recurrence theorem for dependent processes with applications to data compression, A.B. Nobel and A.D. Wyner, *IEEE Transactions on Information Theory*, vol. 38, pp. 1561-1564, 1992.

A note on uniform laws of averages for dependent processes, A.B. Nobel and A. Dembo, *Statistics and Probability Letters*, vol. 17, pp.169-172, 1993.

A counterexample concerning uniform ergodic theorems for a class of functions, A.B. Nobel, *Statistics and Probability Letters*, vol. 24, pp.165-168, 1995.

Termination and continuity of greedy growing for tree-structured vector quantizers, A.B. Nobel and R.A. Olshen, *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp.191-205, 1996.

Consistency of data-driven histogram methods for density estimation and classification, G. Lugosi and A.B. Nobel, *Annals of Statistics*, vol. 24, no. 2, pp.687-706, 1996.

Vanishing distortion and shrinking cells, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp.1303-1305, 1996.

Histogram regression estimation using data-dependent partitions, A.B. Nobel, *Annals of Statistics*, vol. 24, no. 3, pp.1084-1105, 1996.

Recursive partitioning to reduce distortion, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 43, no. 4, pp.1122-1133, 1997.

Density estimation from an individual numerical sequence, A.B. Nobel, G. Morvai and S. Kulkarni, *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp.537-541, 1998.

On density estimation from an ergodic process, T.M. Adams and A.B. Nobel, *Annals of Probability*, vol. 26, no. 2, pp.794-804, 1998.

Limits to classification and regression estimation from ergodic processes, A.B. Nobel, *Annals of Statistics*, vol. 27, pp.262-273, 1999.

Adaptive model selection using empirical complexities, G. Lugosi and A.B. Nobel, *Annals of Statistics*, vol. 27, pp.1830-1864, 1999.

Regression estimation from an individual stable sequence, G. Morvai, S.R. Kulkarni, and A.B. Nobel, *Statistics*, vol. 33, pp.99-118, 1999.

Finitary reconstruction of a measure preserving transformation, T.M. Adams and A.B. Nobel, *Israel Journal of Mathematics*, vol. 126, pp.309-326, 2001.

Estimating a function from ergodic samples with additive noise, A.B. Nobel and T.M. Adams, *IEEE Transactions on Information Theory*, vol. 47, pp.2895-2902, 2001.

Analysis of a complexity based pruning method for classification trees, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 48, pp.2362-2368, 2002.

On optimal sequential prediction schemes for general processes, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 49, pp.83-98, 2003.

Indistinguishability of absolutely continuous and singular distributions, S.P. Lalley and A.B. Nobel, *Statistics and Probability Letters*, vol. 62, pp.145-154, 2003.

Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets, T. Sørlie, R. Tibshirani, J. Parker, T. Hastie, J.S. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, C.M. Perou, P.E. Lønning, P.O. Brown, A-L. Børresen-Dale and D. Botstein, *Proceedings of the US National Academy of Sciences*, vol. 100, pp.8418-8423, 2003.

Some statistical properties of memoryless individual sequences, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 50, pp.1497-1505, 2004.

Significance analysis of functional categories in gene expression studies: a structured permutation approach, W.T. Barry, A.B. Nobel and F.A. Wright, *Bioinformatics*, vol. 21, pp.1943-1949, 2005.

ChIPOTle: A user-friendly tool for the analysis of ChIP-chip data, M.J. Buck, A.B. Nobel and J.D. Lieb, *Genome Biology*, vol. 6, R97, 2005.

Hypothesis testing for families of dependent processes, A.B. Nobel, *Bernoulli*, vol. 12, pp.251-269, 2006.

The Molecular Portraits of Breast Tumors Are Conserved Across Microarray Platforms, Z. Hu, C. Fan, D.S. Oh, J.S. Marron, X. He, B.F. Qaqish, C. Livasy, L.A. Carey, E. Reynolds, L. Dressler, A. Nobel, J. Parker, M.G. Ewend, L.R. Sawyer, D. Xiang, J. Wu, Y. Liu, R. Nanda, M. Tretiakova, A.R. Orrico, D. Dreher, J.P. Palazzo, L. Perreard, E. Nelson, M. Mone, H. Hansen, M. Mullins, J.F. Quackenbush, O.I. Olopade, P.S. Bernard and C.M. Perou, *BMC Genomics*, 7:96, 2006.

Different gene expression-based predictors for breast cancer patients are concordant, C. Fan, D.S. Oh, L. Wessels, A. Nobel, L.J. vant Veer, and C.M. Perou, *The New England Journal of Medicine*, vol. 355, pp.560-569, 2006.

Denoising deterministic time series, S.P. Lalley and A.B. Nobel, *Dynamics of Partial Differential Equations*, vol. 3, pp.259-279, 2006.

Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer, T. Sørlie, C.M. Perou, C. Fan, S. Geisler, T. Aas, A. Nobel, G. Anker, L.A. Akslen, D. Botstein, A-L. Børresen-Dale, and P.E. Lønning, *Molecular Cancer Therapeutics*, vol. 5, pp.2914-2918, 2006.

A statistical framework for testing functional categories in microarray data, W.T. Barry, A.B. Nobel and F.A. Wright, *The Annals of Applied Statistics*, vol. 2, pp.286-315, 2008.

Sequential procedures for aggregating arbitrary estimators of a conditional mean, F. Bunea and A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 54, pp.1725-1735, 2008.

Merging two gene expression studies via cross platform normalization, A.A. Shabalin, H. Tjelmeland, C. Fan, C.M. Perou and A.B. Nobel, *Bioinformatics*, vol. 24, pp.1154-1160, 2008.

Statistical significance of clustering for high dimension low sample size data, Y. Liu, D.N. Hayes, A.B. Nobel and J.S. Marron, *Journal of the American Statistical Association*, vol. 103, pp.1281-1293, 2008.

On the Size and Recovery of Submatrices of Ones in a Random Binary Matrix, X. Sun and A.B. Nobel, *Journal of Machine Learning Research*, vol. 9, pp.2431-2453, 2008.

FastMap: Fast eQTL Mapping in homozygous populations, D.M. Gatti, A.A. Shabalin, T-C. Lam, F.A. Wright, I. Rusyn and A.B. Nobel, *Bioinformatics*, vol. 25, pp.482-489, 2008.

Supervised risk predictor of breast cancer based on intrinsic subtypes, J.S. Parker, M. Mullins, M.C.U. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, J.F. Quackenbush, I.J. Stijleman, J. Palazzo, J.S. Marron, A.B. Nobel, E. Mardis, T.O. Nielsen, M.J. Ellis, C.M. Perou, P.S. Bernard, *Journal of Clinical Oncology*, JCO.2008.18.1370, 2009.

A Bayesian model for cross-study analysis of differential gene expression, R.B. Scharpf, H. Tjelmeland, G. Parmigiani and A.B. Nobel. To appear in the *Journal of the American Statistical Association*, 2009.

Finding large average submatrices in high dimensional data, A.A. Shabalin, V.J. Weigman, C.M. Perou and A.B. Nobel. To appear in the *Annals of Applied Statistics*.

B. Papers Under Review or in Progress, and Technical Reports

Uniform convergence of VC classes under ergodic sampling, T.M. Adams and A.B. Nobel. Submitted.

On the size of large-average and ANOVA-fit submatrices in Gaussian random matrices, X. Sun and A.B. Nobel. In progress.

First order predictive sequences and induced transformations, A.B. Nobel, Technical Report #2367, Department of Statistics, UNC Chapel Hill, 1999.

C. Refereed Conference Papers

Evaluating the performance of a simple inductive procedure in the presence of overfitting error, A.B. Nobel. In *Proceedings of the Fourth Annual Conference on Computational Learning Theory*, pp.267-274, Santa Cruz, CA, 1991.

Understanding Patterns of TCP Connection Usage with Statistical Clustering, F. Hernandez Campos, A.B. Nobel, F.D. Smith, K. Jeffay. To appear in the proceedings of the *Thirteenth IEEE/ACM International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, Atlanta, GA, September 2005.

Mining Approximate Frequent Itemsets from Noisy Data, J. Liu, S. Paulsen, W. Wang, A.B. Nobel and J. Prins (short paper). *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM)*, Houston, TX, November, 2005.

Mining approximate frequent itemsets in the presence of noise: Algorithms and analysis, J. Liu, S. Paulsen, X. Sun, W. Wang, A.B. Nobel and J. Prins. To appear in the *Proceedings of the 2006 SIAM Conference on Data Mining (SDM)*, Bethesda, MD, April 2006.

Significance and Recovery of Block Structures in Binary Matrices with Noise, X. Sun and A.B. Nobel, *Proceedings of the 19th Annual Conference on Learning Theory (COLT)*, H.U. Simon and G. Lugosi eds., Springer, 2006.

D. Patents and Patent Applications

Methods, systems and computer program products for modeling and simulating application-level traffic characteristics in a network based on transport and network layer header information, Traffic, F. Hernandez-Campos, K. Jeffay, F.D. Smith, A.B. Nobel, U.S. Letters Patent No. 7,447,209.

E. Refereed Conference Posters and Short Papers

Multivariate SVD Analyses For Network Anomaly Detection, J. Terrell, L. Zhang, K. Jeffay, A.B. Nobel, H. Shen, F.D. Smith, Z. Zhu, ACM SIGCOMM 2005 Poster Session, Philadelphia, PA, August 2005.

A Non-Parametric Approach to Generation and Validation of Synthetic Network Traffic, F. Hernandez-Campos, A.B. Nobel, F.D. Smith, K. Jeffay, IMA Workshop on Measurement, Modeling, and Analysis of the Internet, Poster Session, Minneapolis, MN, January 2004.

Some statistical properties of memoryless individual sequences, A.B. Nobel, 2004 IEEE International Symposium on Information Theory, Chicago, IL, July 2005.

F. Conference Papers

Histogram density estimation using data-dependent partitions, A.B. Nobel and G. Lugosi. In *Proceedings of the 1994 Conference on Information Science and Systems*, pp.775-780, Princeton University, Princeton, NJ, 1994.

Consistent estimation of a dynamical map, A.B. Nobel. In *Nonlinear Dynamics and Statistics*, pp.267-280, edited by A.I. Mees, Birkhauser, Boston, 2001.

PRESENTATIONS

A. Invited Presentations at Conferences

"Histogram density estimation using data-dependent partitions", *Conference on Information Science and Systems*, Princeton University, Princeton, NJ, 1994.

"Greedy growing of tree-structured classification rules using a composite splitting criterion", *IEEE-IMS Workshop on Information Theory and Statistics*, Alexandria, VA, 1994.

"Analysis of several recursive growing algorithms for classification quantization trees", *IMS/WNAR Meeting*, Stanford University, 1995.

"Adaptive model selection using empirical complexities", presented by G. Lugosi, *4th World Congress of the Bernoulli Society*, Vienna, Austria, 1996.

"Adaptive model selection using empirical complexities", *Non-Gaussian Signal Processing Workshop*, Fort Meade, MD, 1996.

"Limits to nonparametric estimation from ergodic processes, and some positive results", *Workshop on Non-Linear Time Series for Prediction and Control*, Technion, Haifa, Israel, June 1998.

"Estimating a measure preserving transformation from a single trajectory", *Workshop on Nonlinear Dynamics and Statistics*, Isaac Newton Institute, Cambridge University, England, 1998, and *Joint Statistical Meetings*, Baltimore, MD, August 1999.

"Transformations and singular sequences", *936'th AMS Sectional Meeting*, Winston-Salem, NC, 1998.

"Classification from ergodic samples and individual sequences", *IEEE Information Theory Workshop on Detection, Estimation, Classification, and Imaging*, Santa Fe, NM, 1999.

"Denoising deterministic time series", *International Conference on Stochastic Optimization and Adaptation*, Cochin, India, 2000.

"The role of cluster analysis in the study of gene expression arrays", *National Science Foundation Pattern Recognition Workshop*, University of Michigan, Ann Arbor, MI, 2002.

"Denoising deterministic time series and some related questions of statistical inference", *Lucent Distinguished Lecture Series in Communications and Networks*, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, 2002.

"Distinguishing families of dependent processes", presented at the *Nonparametric Statistics Research Conference*, Florida State University, Tallahassee, FL, 2003, and the *Ergodic Theory Workshop*, UNC Chapel Hill, NC, 2003, and the *2003 Joint Statistical Meetings*, San Francisco, CA, 2003.

"Statistical Clustering of Internet Communication Patterns", presented by F. Hernandez-Campos, *35th Symposium on the Interface: Computing Science and Statistics*, Salt Lake City, Utah, 2003.

"Significance Analysis of Bi-clustering for Gene Expression Data", *New Inference Concepts for Analysing Complex Data*, Mathematisches Forschungsinstitut Oberwolfach, 2004, and *Second IMS-ISBA Joint Meeting*, Bormio, Italy, 2005.

"Significance and Recovery of Block Structures in Binary Matrices with Noise", workshop on *Statistische und Probabilistische Methoden der Modellwahl*, Mathematisches Forschungsinstitut Oberwolfach, 2005.

"A Bayesian Model for Cross-Study Differential Expression", ASA Session on Statistical Methods for Integrative Genomics, *Joint Statistical Meetings*, 2006.

"Finding significant large-average blocks in high dimensional data", *39th Symposium on the Interface: Computing Science and Statistics*, Philadelphia, PA, 2007.

Invited participant, *NSF Workshop on Discovery in Massive or Complex Data Sets*, Washington, DC, October, 2007.

B. Conference Abstracts

S.H. Randell, S. Gabriel, S. Deng, A. Nobel and W. O'Neal, Novel pro- and anti-inflammatory mechanisms revealed by mRNA expression analysis of Pseudomonas aeruginosa (Ps.a.)-treated primary human tracheobronchial epithelial (HTBE) cells. *Pediatric Pulmonology Supplement* 24:241, 2002.

C. Contributed Presentations at Conferences

"Distributed information storage". Presented by J. Roche at the *IEEE International Symposium on Information Theory*, Kobe, Japan, June 1988.

"Some closure properties for uniformly convergent classes of functions". Presented at the *Twentieth Conference on Stochastic Processes and their Applications*, Nahariya, Israel, June 1991.

"On uniform convergence for dependent processes". Presented at the *IEEE International Symposium on Information Theory*, Budapest, Hungary, June 1991.

"Evaluating the performance of a simple inductive procedure in the presence of overfitting error". Presented at the *Fourth Annual Conference on Computational Learning Theory*, Santa Cruz, CA, August 1991.

"Boundedness and consistency of greedy growing for tree-structured vector quantizers". Presented at the *IEEE International Symposium on Information Theory*, San Antonio, TX, January 1993.

"Histogram classification using vector quantization". Presented at the *IEEE International Symposium on Information Theory*, Trondheim, Norway, June 1994.

"Continuity and consistency of greedy growing for tree-structured vector quantizers". Presented at the *IEEE International Symposium on Information Theory*, Trondheim, Norway, June 1994.

"Histogram regression estimation using data-dependent partitions". Presented at the *1995 IMS Annual Meeting*, Montreal, Quebec, July 1995.

"Vanishing distortion and shrinking cells". Presented at the *1996 IEEE Information Theory Workshop*, Haifa, Israel.

"An adaptive model selection procedure using empirical complexities". Presented at the workshop on *Adaptive Selection of Models and Statistical Procedures*, Mt. Holyoke College, South Hadley, MA, June 1996.

"Consistency of empirically optimal regression trees". Presented at the *1996 Joint Statistical Meetings, Annual Meeting of the IMS*, Chicago, Il, August 1996.

"Some limits to nonparametric estimation from ergodic time series". Presented at the *Symposium on Nonparametric Functional Estimation*, Centre de recherches mathématiques, Université de Montréal, Montreal, Canada, October 1997.

"Density estimation from an individual numerical sequence". Presented at the conference on *Recent Advances in Statistics and Probability*, Indian Statistical Institute, Calcutta, India, December 1997, and the *1998 IEEE International Symposium on Information Theory*, MIT, Cambridge, MA, August 1998.

"Denoising deterministic time series". Presented at the conference *Information Theory in Mathematics*, Balatonlelle, Hungary, July 2000, and the *MSRI Workshop on Nonlinear Estimation and Classification*, Berkeley, CA, March 2001.