Short CV
James G. Booth

April 2019

Contact information

Department of Statistics and Data Science, 1172 Comstock Hall, Cornell University, Ithaca, NY 14853. Email: Jim.Booth@cornell.edu

Profile

I am currently a Professor in the Department of Statistics and Data Science (SDS) at Cornell University, one of three departments in Computing and Information Science. After visiting the Department of Operations Research and Information Engineering at Cornell in 2003 I was hired in the Department of Biological Statistics and Computational Biology, in the College of Agricultural and Life Sciences. From 1987 to 2003 I was a faculty member in the Department of Statistics at the University of Florida. During that period I spent two years as a Research Fellow at the Australian National University, and one year at Colorado State University.

My research interests involve basic statistical methodology including: the bootstrap and Monte Carlo methods, clustering, exact inference, mixed models, generalized linear models, and also applications in bioinformatics. I have taught a variety of courses at Cornell including Statistical Methods II, the second semester of a statistical methods sequence for graduate students from a wide variety of disciplines; Biological Statistics I, an introductory statistics course and a core requirement for undergraduate statistics majors; Data Science for All, based on the data8 course developed at Berkeley; as well as core courses for statistics undergraduates, professional masters students, and Ph.D. students in the Fields of Statistics. As a CALS faculty member in SDS part of my teaching effort involves contributions to the campus-wide statistical consulting service through the Cornell Statistical Consulting Unit.

Education

PhD	Statistics	1987	University of Kentucky
MS	Statistics	1984	University of Kentucky
MSc	Statistics	1982	University of Leeds, England
BSc	Mathematics	1981	University of Leeds, England

Positions Held

Professor	2004 –present	Cornell University
Visiting Professor	Fall 2016	University of Connecticut
Visiting Professor	2003 – 2004	Cornell University
Professor	2000 – 2004	University of Florida
Visiting Professor	1996 – 1997	Colorado State University
Associate Professor	1993 – 2000	University of Florida
Research Fellow	1990 – 1992	Australian National University
Assistant Professor	1987 – 1993	University of Florida
Instructor	1986 – 1987	Transylvania University, KY

Departmental Service

- *Director of Graduate Studies*, Field of Statistics, Cornell University, 2017–present.
- *Chair*, Department of Biological Statistics and Computational Biology, College of Agriculture and Life Sciences, Cornell University, 2007–2016
- *Director of Undergraduate Studies*, Biometry and Statistics, College of Agriculture and Life Sciences, Cornell University, 2011–2016
- *Acting Chair*, Department of Statistics, University of Florida, Fall 2002
- *Associate Chair*, Department of Statistics, University of Florida, 2001–2003
- *Graduate Coordinator*, Department of Statistics, University of Florida, 1995–1996, 1997–2001

Professional Service

- *External Review Committee*: Department of Statistics, Southern Methodist University, Nov. 2012.
- *Executive Committee*: Statistical Modelling Society, 2006-2012.
- *Associate Editor*: Journal of Statistical Planning and Inference, 2012–2014, Journal of the American Statistical Association, 2006–2012, Statistical Modelling: 2001–, ANZ Journal of Statistics: 2001–2008, Journal of the Royal Statistical Society, Series B: 2001–2005, Computational Statistics: 1996–2002
- *Book Review Board*: Journal of the American Statistical Association: 1999–2002
- *NSF Review Panel*: Program in Statistics: Dec. 10-12, 2001 and Feb. 6-8, 2013.

Graduate Students Supervised

- Kelson Zawack, PhD in Computational Biology, Cornell University, May 2017 (co-adviser: Yrjo Grohn). Postdoctoral Assocate, Department of Biostatistics, Yale University.
- Irina Gaynanova, PhD, Cornell University, May 2015 (co-adviser: Martin Wells). Assistant Professor, Department of Statistics, Texas A&M University.
- Muting Wan, PhD, Cornell University, January 2015. Data Scientist, New York Life.
- Caitlin Cunningham, Cornell University, August 2012. Assistant Professor, LeMoyne College, Syracuse.
- Haim Bar, PhD, Cornell University, January 2012. Assistant Professor, Department of Statistics, University of Connecticut.
- Kirsten Eilertson, PhD, Cornell University, August 2011 (Co-adviser: Carlos Bustamante). Research Associate, Department of Statistics, Pennsylvania State University.
- Matthias Kormacksson, PhD, Cornell University, August 2009. Research Scientist, Novartis.
- Vadim Zipunikov, PhD, Cornell University, January 2008. Assistant Professor, Department of Biostatistics, Johns-Hopkins University.
- David Hitchcock, PhD, August 2004 (co-advisor: George Casella). Associate Professor at University of South Carolina.
- Bernhard Klingenberg, PhD, August 2004 (co-advisor: Alan Agresti). Associate Professor, Williams College.

- Brian Caffo, PhD, August 2001. Professor, Department of Biostatistics, Johns-Hopkins University.
- Wolfgang Jank, PhD, August 2001. Anderson Professor of Global Management, Department of Information Systems and Decision Sciences, University of South Florida.
- Glen Hartless, PhD, August 2000 (co-advisor: Ramon Littell). CIA Analyst in Washington D.C.
- Somnath Sarkar, PhD, August 1996 (co-advisor: M. Ghosh). Statistician at Eli-Lilly, Indianapolis.

Grants

- co-PI (PI Haim Bar, co-PI Martin Wells), NSF DMS-1611893, "Variable selection when p>>N - beyond the linear regression and normal errors model" 2016-2019.
- PI (co-PI Martin Wells), NSF DMS-1208488, "Models and computational strategies in statistical bioinformatics" 2012-2014.
- PI, NSA Grant 08C-016, "24th International Workshop on Statistical Modeling" 2008-2009.
- PI, NSF DMS-0805865, "Applications and computational issues involving generalized linear and mixed models", 2008-2010.
- Co-PI, NSF Grant DMS-0405543, "Cluster analysis, predictive distributions, and stochastic search algorithms," 2004-2008. PI George Casella, Department of Statistics, University of Florida.
- Co-PI, NSF Grant DMS-0072827, "Combining EM and Monte Carlo to maximize intractable likelihood functions," 2000-2004.
- PI, NSF Grant DMS-9813374, "NSF/CBMS Regional Conference in the Mathematical Sciences - Generalized Linear Mixed Models and Related Topics - June 8-12, 1999."
- PI, NSF Grant DMS-93010836, "Some new bootstrap methods for sample surveys," 1993-1996.

Selected Publications

[1] Bar, H., Booth, J. G., Liu, K. & Wells, M. T. (2018). Facilitating high-dimensional transparent classificiation via empirical bayes variable selection. *Applied Stochastics Models in Business and Industry* Published as Early View.

[2] Bar, H. Y., Booth, J. G. & Wells, M. T. (2014). A bivariate model for simulataneous testing in bioinformatics data. *Journal of the American Statistical Association* **109**, 537–547.

[3] Booth, J. G. & Butler, R. W. (1999). An importance sampling algorithm for exact conditional tests in log-linear models. *Biometrika* **86**, 321–332.

[4] Booth, J. G., Butler, R. W. & Hall, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association* **89**, 1282–1289.

[5] Booth, J. G., Capanu, M. & Heigenhauser, L. (2005). Exact conditional p-value calculation for the quasi-symmetry model. *Journal of Computational and Graphical Statistics* **14**, 716–725.

[6] Booth, J. G., Casella, G. & Hobert, J. P. (2008). Clustering using objective functions and stochastic search. *Journal of the Royal Statistical Society* **B 70**, 119–139.

[7] Booth, J. G., Federer, W. T., Wells, M. T. & Wolfinger, R. (2009). A multivariate variance component model for analysis of covariance in designed experiments. *Statistical Science* **24**, 223–237.

[8] Booth, J. G. & Hall, P. (1994). Monte Carlo approximation and the iterated bootstrap. *Biometrika* **81**, 331–340.

[9] Booth, J. G., Hall, P. & Wood, A. T. A. (1993). Balanced importance resampling for the bootstrap. *Annals of Statistics* **21**, 286–298.

[10] Booth, J. G. & Hobert, J. P. (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association* **93**, 262–272.

[11] Booth, J. G. & Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society* **B 61**, 265–285.

[12] Booth, J. G. & Presnell, B. (1998). Allocation of Monte Carlo resources for the iterated bootstrap. *Journal of Computational and Graphical Statistics* **7**, 92–112.

[13] Caffo, B. S. & Booth, J. G. (2001). A Markov chain Monte Carlo algorithm for approximating exact conditional tests. *Journal of Computational and Graphical Statistics* **10**, 730–745.

[14] Caffo, B. S. & Booth, J. G. (2003). Monte Carlo conditional tests for log-linear and logistic models: a survey of current methodology. *Statistical Methods in Medical Research* **12**, 1–15.

[15] Caffo, B. S., Booth, J. G. & Davison, A. C. (2002). Empirical sup rejection sampling. *Biometrika* **89**, 745–754.

[16] Gaynanova, I., Booth, J. G. & Wells, M. T. (2016). Simultaneous sparse estimation of canonical vectors in the p>>n setting. *Journal of the American Statistical Association* **111**, 696–706. Published online 16 Apr 2015.

[17] Gaynanova, I., Booth, J. G. & Wells, M. T. (2017). Penalized versus constrained generalized eigenvalue problems. *Journal of Computational and Graphical Statistics* **26**, 379–387. Posted online 06 Apr 2016.

[18] Hartless, G. L., Booth, J. G. & Littell, R. C. (2003). Local influence of predictors in multiple linear regression. *Technometrics* **45**, 326–332.

[19] Hitchcock, D. B., Casella, G. & Booth, J. G. (2006). Improved estimation of dissimilarities by smoothing functional data. *Journal of the American Statistical Association* **101**, 211–222.

[20] Jank, W. & Booth, J. G. (2003). Efficiency of Monte Carlo EM and simulated maximum likelihood in generalized linear mixed models. *Journal of Computational and Graphical Statistics* **12**, 214–229.

[21] Kormaksson, M., Booth, J. G., Figueroa, M. E. & Melnick, A. (2012). Integrative model-based clustering of microarray methylation and expression data. *Annals of Applied Statistics* **6**, 1327–1347.