These multivariate spacings can be viewed as a data-driven realization of the so-called "statistically equivalent blocks". These spacings assume a form of center-outward layers of "shells" ("rings" in the two-dimensional case), for which the shapes of the shells follow closely the underlying probabilistic geometry. We discuss the properties and applications of these spacings. For example, we use the spacings to construct tolerance regions, and show their applications to monitoring aircraft landings. The construction is nonparametric and completely data driven, and the resulting tolerance region reflects the true geometry of the underlying distribution. This is different from the existing approaches which require that the shape of the tolerance region be specified in advance.

Finally, we also discuss several families of multivariate goodness-of-fit tests based on the proposed multivariate spacings.

DD-CLASSIFIERS: NEW NONPARAMETRIC CLASSIFICATION PROCEDURE

Wednesday, April 7, 2010, 10:30-11:30 a.m.
Hawking Auditorium, Mitchell Physics Building

Most existing classification algorithms are developed by assuming either certain parametric distributions for the data or certain forms of separating surfaces. Either assumption can greatly limit the potential applicability of the algorithm. We introduce a new nonparametric classification algorithm using the so-called DD-plot. This algorithm is completely nonparametric, requiring no prior knowledge of the underlying distributions or of the form of the separating surface. Thus it can potentially be applied to a wide range of classification problems. The algorithm can be easily implemented and its classification outcome can be clearly visualized on a two-dimensional plot regardless of the dimension of the data. Furthermore, it can also be robust against outliers or contamination. The asymptotic properties of the proposed classifier and its misclassification rate are discussed. The DD-classifier is shown to be asymptotically equivalent to the Bayes rule under suitable conditions. The performance of the classifier is also examined by using simulated and real data sets. Overall, the DD-classifier performs well across a broad range of settings, and compares favorably with most existing nonparametric classifiers.

Herman Otto Hartley, Distinguished Professor

The Texas A&M University Department of Statistics established the H. O. Hartley Memorial Lectures in 1988 to honor the memory of Prof. **Herman Otto Hartley.** Hartley accepted an appointment as Distinguished Professor at Texas A&M in 1963, founded the Texas A&M Institute of Statistics, and served as its director until his retirement in 1977. Hartley built his initial faculty of four into a group of 16, directed more than 30 doctoral students, and published over 75 papers during this period. He served as president of the American Statistical Association in 1979. Professor Hartley died on December 30, 1980.

Hartley was well known for his work on the foundations of sampling theory, and made important contributions to mathematical optimization, estimation with incomplete data, estimation of variance components, and establishment of safe doses in carcinogenic experiments. Hartley collaborated with Egon Pearson to produce the classic two-volume *Biometrika Tables for Statisticians.*

Hartley earned a Ph.D. degree in mathematics at Berlin University in 1934 and a Ph.D. degree in statistics under John Wishart at Cambridge University in 1940. He taught at University College, London and Iowa State College before coming to Texas A&M. Hartley was deeply committed to all phases of his profession, including education, research, and delivery of knowledge and advice to users of statistics. H. O. Hartley was not only a brilliant academician, but also a warm and caring human being. His legacy continues to have a profound influence on Texas A&M statistics and the larger statistical community.

The H. O. Hartley Memorial Lectures are held every other year and were first given in 1988 by Peter J. Diggle. Subsequent lecturers have been Bradley Efron, E. J. Hannan, Sir David R. Cox, Wayne A. Fuller, Adrian Raftery, Peter Hall, Terry Speed, Jim Berger, and Edward Goerge. This year the Department of Statistics is honored to have Prof. **Regina Liu** present the H. O. Hartley Lectures.

Regina Liu, Professor and Chair
Department of Statistics & Biostatistics, Rutgers University

Regina Liu received her Ph.D. in Statistics from Columbia University in 1983 under the direction of John Van Ryzin. She was named full professor in the Department of Statistics and Biostatistics at Rutgers University in 1994 and was then named Chair of Statistics and Biostatistics in 2005.

Liu was the Editor-in-Chief of the *Journal of Multivariate Analysis* from 1993-1998 and has been an active Associate Editor for journals such as the *Annals of Statistics* and the *Journal of the American Statistical Association*. She has written three books and has written software for online monitoring of manufacturing or inspection data using bootstrap method.

Among the awards and honors that Dr. Regina Liu has received are Elected Fellow of the American Statistical Association, Elected Fellow of the Institute of Mathematical Statistics, and the Fulbright Scholar from Spain in 2008. She also received the Best Paper Award for Feature Application for 2000-2001 from the Institute of Industrial Engineers. Dr. Liu has published more than 50 papers and has been invited to nearly 100 talks throughout her distinguished career.

Dr. Liu's research interests include data quality, text mining, nonparametric inferences, Resampling, applications of data depth, and aviation safety analysis.

STATISTICS IS A MANY-SPLENDORED THING: MINING MASSIVE TEXT DATA AND BEYOND

Monday, April 5, 2010, 4:00-5:00 p.m.
Hawking Auditorium, Mitchell Physics Building

Advances in powerful data acquisition technologies have made the collection of massive data a routine practice in all fields. Massive data, ranging from numerical to imaging or textual, are ubiquitous nowadays. Statisticians play a role of ever-increasing importance in such an era of data explosion. They are expected to "make the data tell a story". This can be done in many contexts and many ways, so that statistics is truly a many-splendored thing!

Working with the massive FAA aviation safety report repository, we have developed a systematic data mining procedure for exploring large free-style text datasets. The procedure helps discover useful features for tracking aviation performance. We will discuss some specific text analysis methodologies and tracking statistics. Finally, we apply this data mining procedure to analyze the FAA report repository, illustrate its utility and tell a story in aviation risk management.

Although most illustrations here are drawn from aviation safety data, the proposed data mining procedure applies to many other domains, including, for example, mining free-style medical reports for tracking possible disease outbreaks.

DATA DEPTH FOR MULTIVARIATE SPACINGS, ORDERING, TOLERANCE REGIONS, etc.

Tuesday, April 6, 2010, 10:30-11:30 a.m.
Hawking Auditorium, Mitchell Physics Building

There has been considerable renewed interest in nonparametric multivariate statistics due to the development of data depth and its induced center-outward ordering (or ranking) of multivariate data. We highlight some of the recent advances. In particular, we introduce and develop multivariate spacings using the order statistics derived from data depth. Specifically, the spacing between two consecutive order statistics is the region which bridges the two order statistics, in the sense that the region contains all the points whose depth values fall between the depth values of the two consecutive order statistics.