VOLUME 17, NO 1, JUNE 2006

A joint newsletter of the Statistical Computing & Statistical Graphics Sections of the American Statistical Association

Statistical COMPUTING & GRAPHICS

A Word from our 2006 Section Chairs

PAUL MURRELL
GRAPHICS

I would like to begin by highlighting a couple of interesting recent developments in the area of Statistical Graphics.

There has been a lot of activity on the GGobi project lately, with an updated web site, new versions, and improved links to R. I encourage you to (re)visit www.ggobi.org and see what they've been up to.

The third volume of the Handbook of Computational Statistics, which is focused on Data Visualization, is scheduled for publication at the end of this year and there will be a workshop as a satellite of Compstat 2006. This important volume will contain over 30 contributions and will provide a comprehensive overview of all areas of data visualization. Information about this project is available at gap.stat.sinica.edu.tw/HBCSC.

The big event for our section is of course the JSM in Seattle. Our program chair Juergen Symanzik and our program chair-elect Simon Urbanek have organised a great program of invited sessions, contributed sessions, and roundtables. Juergen provides more details later in this newsletter.

Continues on Page 2.....

STEPHAN R. SAIN
COMPUTING

When Carey Priebe asked me to run for one of the section offices a couple of years ago, I wasn't exactly sure what I was getting into. Now that I have been chair for a couple of months, I'm still not totally sure what I've gotten into. The one thing I do know, though, is that I'm very happy to be involved in the section. There are a lot of very interesting things going on and, as always, a lot of opportunity for people who have an interest in statistical computing.

Continues on Page 2.....

Graphics, Continues from Page 1....

The section is also sponsoring a continuing education course, "Creating More Effective Graphs", given by Naomi Robbins (one of our own Council of Sections reps), and we are sponsoring a special poster session for the Data Expo competition www.amstat-online.org/sections/graphics/dataexpo/ From what I have seen of the Data Expo entries so far this will be a very interesting and stimulating session. There will also be our traditional joint mixer with the Statistical Computing section, which is always a lively event! Visit the on-line program to see all that we have in store www.amstat.org/meetings/jsm/2006/.

Winter is sinking its wet and freezing claws into New Zealand right now, so I'm looking forward to a Summer visit to Seattle! But the weather down here will be much nicer by next February, which is when the Department of Statistics at The University of Auckland will be hosting the DSC 2007 conference (Directions in Statistical Computing) www.stat.auckland.ac.nz/dsc-2007/. If you have never been to New Zealand, this might be a good opportunity to give it a try!

Finally, looking even further ahead, our program chair-elect, Simon Urbanek will be on the lookout for ideas for invited sessions for JSM 2007. Please contact him if you have any good suggestions.

Computing, Continues from Page 1....

The program for the year's JSM in Seattle, WA, is very exciting. We have eight invited sessions, ranging across a wide array of topics including clustering, climate, human-genome studies, social relationships, massive datasets, machine learning, and Monte Carlo methods. In addition, there is an invited session in memory of Leo Breiman, also focusing on machine learning. Another session not to be missed is the topic contributed session where the student award winners will be speaking.

In addition to the fantastic slate of talks and posters at the JSM, the section is sponsoring two continuing education courses. Both of which are highly relevant and focus on topics of great interest and importance.

The first course is on text mining and is put together by David Madigan of Rutgers University and David D. Lewis of David D. Lewis Consulting LLC. The second centers on Monte Carlo integration and optimization and will be conducted by Jennifer Hoeting and Geof Givens from Colorado State University. I am also looking forward to seeing the results of the Data Expo, sponsored by the Section on Statistical Graphics along with the Section on Statistical Computing and Statistics and the Environment. The topic is very interesting, focusing on a graphical analysis of geographic and atmospheric variables. Entries can be viewed in a special topic contributed poster session.

In addition to the JSM this year, the section helped sponsor a workshop on Fast Manifold Learning held in April at the College of William \& Mary. Put together by Michael Trosset and Carey Priebe, the workshop appeared to be a great success and brought together a fantastic and diverse group of people to discuss a topic of increasing interest to the statistical computing community.

The section is sponsoring a two travel scholarships (including one targeted for a student) for the Bioconductor conference just prior to the JSM this August. Also, the section is helping support the useR! conference in Vienna on June 15-17. These are important activities for software developers and researchers with an interest in statistical computing. I would like to take this opportunity to thank the current and past officers of the section for the outstanding job that they are doing and have done with the section. In particular, I would like to thank past-chairs Tim Hesterberg and Carey Priebe, who both have been very helpful and fantastic resources for me, and chair-elect John Monahan. Others include Mike Trosset, the current Program Chair who has put together a great slate for the JSM, Program Chair-Elect Ed Wegman, Treasurer David Poole, and Council of Section Representatives Juana Sanchez, Vincent Carey, and Robert Gentleman.

I would also like to acknowledge Juana Sanchez, our Newsletter Editor, who has been so diligent (and incredibly patient) with putting this newsletter together. Also, Electronic Communication Liaison Thomas Devlin, Continuing Education Liaison John Miller, Awards Officer Jose Pinheiro (who has done an incredible job during his tour as awards officer), and Todd Ogden, our Publication Officer who also has taken on the task of Webmaster.

Finally, I think it is clear that statistical computing is a very active field and is booming with the current

climate of increasingly accessible computing resources and the every increasing numbers of large, complex, and interesting data problems. The section is constantly on the look out for ways to continue to support the efforts of those involved in statistical computing. If anybody has any ideas are suggestions, in particular ideas involving the support and development of students, please let me know.

Editorial Note

This issue of the newsletter features three articles. Jan De Leeuw tells us about generalizations of the EM algorithm. Michael Lawrence and Duncan Temple Lang describe a new R package for writing GUIs, that is portable across platforms. Ivo Dinov reports on his teaching software, Statistical Online Computational Resource, SOCR. Thanks to these authors for their contributions. It would be helpful to have more contributed articles, so please consider contributing your work.

The News section is loaded with information about the upcoming JSM '06 in Seattle. Michael Trosset details the Statistical Computing program and Juergen Symanzik highlights the Graphics program. Don't forget to attend the Joint Computing and Graphics Mixer on Monday starting at 7:30pm. There are always good conversations, food, drinks and lots of door prizes to win. Summaries of two recent conferences, the Interface between Computing Science and Statistics and Fast Manifold Learning, are provided by Yasmin Said and Michael Trosset, respectively.

Finally, there are two informative articles about the field. Tim Hesterberg reports on computing and graphics activities at Insightful. Robert Gould announces the birth of the new free electronic Journal of Technological Innovations in Statistics Education. An analysis of publications in the existing Journal of Statistical Education, conducted with Nathan Yau, shows a need for a publication outlet for technological research.

Di Cook and Juana Sanchez

Featured Article

SOME MAJORIZATION TECHNIQUES

Jan de Leeuw
University of California, Los Angeles
deleeuw@stat.ucla.edu
http://gifi.stat.ucla.edu

1. Introduction

Majorization algorithms (Deleeuw, 1994; Heiser, 1995, Lange et al., 2000) are used with increasing frequency in statistical computation. They generalize the EM algorithm to a much broader class of problems and they can usually be tailored to handle very high-dimensional problems.

The general idea is simple. If we want to minimize f over $X \subseteq R^n$, we construct a *majorization function* g on X × X such that

$$f(x) \leq g(x,y) \qquad \forall x, y \in X$$
$$f(x) = g(x,y) \qquad \forall x \in X$$

Thus g, considered as a function of x is never below f and touches f at y.

The majorization algorithm corresponding with this majorization function g updates x at iteration k by

$$x^{(k+1)} \in \underset{x \in X}{\mathbf{argmin}}\ g\left(x, x^{(k)}\right),$$

unless we already have

$$x^{(k)} \in \underset{x \in X}{\mathbf{argmin}}\ g\left(x, x^{(k)}\right),$$

in which case we stop. Convergence follows, under some additional simple conditions, from the *sandwich inequality,* which says that if we do not stop at iteration k, then

$$f\left(x^{(k+1)}\right) \leq g\left(x^{(k+1)}, x^{(k)}\right) < g\left(x^{(k)}, x^{(k)}\right) = f\left(x^{k}\right)$$

Consider the example $f(x)=1/4\ x\ +\ 1/2\ x^2$. The function

has local minima at +1 and -1, with function value -1/4, and a local maximum at 0 with function value 0. A majorization function can be constructed by using $x^2 \geq y^2 + 2y(x-y)$, giving $g(x,y)=1/4\ x^4+ 1/2\ y^2\ xy$. This leads to the majorization algorithm $x^{(k+1)} = \sqrt[3]{x^{(k)}}$. This converges linearly, with convergence rate 1/3, to either +1 or -1, depending on where we start.

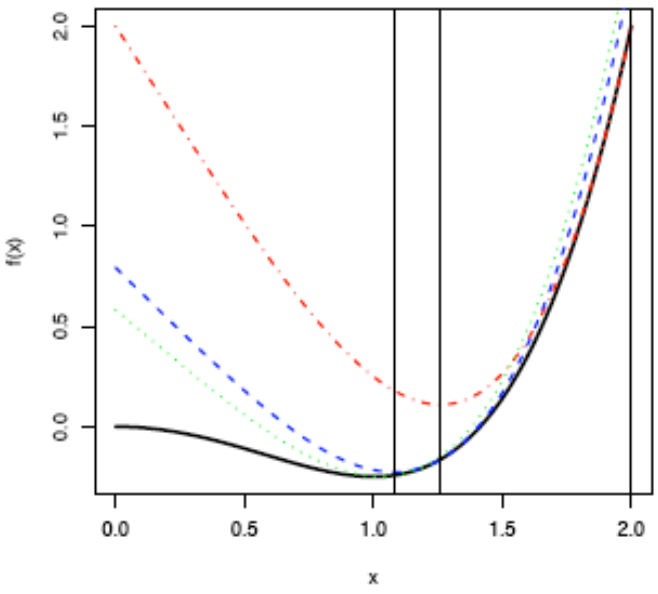


FIGURE 1. Majorization Example

In Figure 1 the function we are minimizing is the thick line.We start at 2. The majorization function at this point, drawn in red with dots-dashes, touches f at 2 and is minimized at $\sqrt[3]{2} \approx 1.2599$. We compute the new majorization function at this point, in blue with dashes, and minimize it at $\sqrt[9]{2} \approx 1.0801$. The next step (green, with dots) takes us to $\sqrt[27]{2} \approx 1.0260$.

Observe that our algorithm is $x^{(k)} = x^{3-k}$. An equally valid majorization algorithm is $x^{(k)} = (-1)^k x^{\frac{1}{3^k}}$, which also minimizes the majorization function in each step. It produces a decreasing sequence of loss function values converging to 1/4, but the sequence of solutions is not convergent and has two converging subsequences, one converging to -1 and one to +1.

In most cases majorization methods converge at a linear rate, with the rate equal to the largest eigenvalue in modulus of the matrix

$$I - \left[D_{11}g(x,x)\right]^{-1} D^2 f(x) = -\left[D_{11}g(x,x)\right]^{-1} D_{12}g(x,x)$$

where the derivatives are evaluated at the fixed point x (Ortega and Rheinboldt, 1970,page 300-301). In some special cases we can have sub-linear or super-linear convergence, but linear convergence is the rule.

2. Using Elementary Inequalities

The first way to construct majorization functions is the simplest one. There are many inequalities in the literature of the form $F(x,y) \geq 0$ with equality if and only if $x=y$. Such inequalities can often be used to construct majorization functions. Since this is not really a systematic approach, we merely illustrate it by a rather detailed example.

After a suitable choice of coordinates and normalization the Euclidean multidimensional scaling problem can be formulated as minimization of

$$(1) \qquad \sigma(x) = 1 + \frac{1}{2}x'x - \sum_{i=1}^{n} w_i \delta_i d_i(x)$$

Here the w_i are known positive *weights,* the δ_i are the *dissimilarities*, and the $d_i(x)$ are the *Euclidean distances*, defined by $d_i(x) = \sqrt{x' A_i x}$. The A_i are known positive semi-definite matrices that satisfy $\sum_{i=1}^{n} w_i A_i = I$.

In most cases of interest the dissimilarities will be positive, but we shall cover the more general case in which there can be both positive and negative ones. Decomposing δ_i into its positive and negative parts, i.e. $\delta_i = \delta_i^+ - \delta_i^-$ with both δ_i^+ and δ_i^- non-negative. Now we can write

$$(2) \qquad \sigma(x) = 1 + \frac{1}{2}x'x - \sum_{i=1}^{n} w_i \delta_i^+ d_i(x) + \sum_{i=1}^{n} w_i \delta_i^- d_i(x)$$

If $d_i(y)>o$ then by, respectively, the Cauchy-Schwartz and the Arithmetic-Geometric Mean Inequality

$$\frac{1}{d_i(y)}x'A_iy \le d_i(x) \le \frac{1}{d_i(y)}\frac{1}{2}\left(x'A_ix + y'A_iy\right),$$

Thus

$$(3a) \qquad \sum_{i=1}^n w_i\delta_i^+ d_i(x) \ge x'B^+(y)y,$$

with

$$(3b) \qquad \mathrm{B}^+(y) = \sum_{i=1}^n w_i \frac{\delta_i^+}{d_i(y)} A_i,$$

And

$$(3c) \qquad \sum_{i=1}^n w_i\delta_i^- d_i(x) \le \frac{1}{2}\left(x'B^-(y)x + y'B^-(y)y\right),$$

with

$$(3d) \qquad \mathrm{B}^-(y) = \sum_{i=1}^n w_i \frac{\delta_i^-}{d_i(y)} A_i,$$

Observe that both B^+ and B^- are positive semi-definite. Combining these results gives

$$(4) \quad \sigma(x) \le 1 + \frac{1}{2}x'x - x'B^+(y)y + \frac{1}{2}x'B^{-1}(y)x + \frac{1}{2}y'B^{-1}(y)y.$$

The right-hand side of (4) gives a quadratic majorization function, and the corresponding algorithm is

$$x^{(k+1)} = \left[I + B^-\left(x^{(k)}\right)\right]^{-1} B^+\left(x^{(k)}\right)x^{(k)}.$$

At a stationary point x the derivative of the algorithmic map is

$$\left[I + B^-(x)\right]^{-1}\left[B^+(x) - H^+(x) + H^-(x)\right],$$

where

$$\mathrm{H}^+(x) = \sum_{i=1}^n w_i \frac{\delta_i^+}{d_i^3(x)} A_ixx'A_i,$$

and

$$\mathrm{H}^-(x) = \sum_{i=1}^n w_i \frac{\delta_i^-}{d_i^3(x)} A_ixx'A_i,$$

The matrices H^+, H^-, and B^+-H^+ are all positive semidefinite.

3. Integrals

Supposed want to maximize

$$f(x) = \log \int_Z \exp\{x,z\}dz$$

Because we are maximizing we will now construct a minorization function and a minorization algorithm.

Of course the logarithm in the definition of f is really irrelevant here and the exponent merely guarantees that we are integrating a positive function. Write

$$f(x) - f(y) = \log \frac{\int_Z \exp\{u(y,z)\}\frac{\exp\{u(x,z)\}}{\exp\{u(y,z)\}}dz}{\int_Z \exp\{u(y,z)\}dz}$$

Jensen's inequality, or equivalently the concavity of the logarithm, tells us that

$$f(x) - f(y) \ge \frac{\int_Z \exp\{u(y,z)\}\log\frac{\exp\{u(x,z)\}}{\exp\{u(y,z)\}}dz}{\int_Z \exp\{u(y,z)\}dz}$$

Define

$$\pi(z \mid y) = \frac{\exp\{u(z,y)\}}{\int_Z \exp\{u(z,y)\}}.$$

Then

$$f(x) \ge f(y) + \int_Z \pi(z \mid y)u(x,z)dz - \int_Z \pi(z \mid y)u(y,z)dz,$$

which defines our minorization function.

A step of the minorization algorithm simply maximizes (in the "M'' step) the ``expectation'' $\int_Z \pi(z \mid y) u(x,z)\,dz.$

Computing, and possibly simplifying, this expectation is the ``E" step. The algorithm is especially attractive, of course, if the integral defining the expectation can be evaluated in closed form. This is often the case in exponential family problems in statistics, where we want to compute maximum likelihood estimates. It is hardly necessary to give an example in this case, because so many examples of the EM algorithmare available.

4. Using Convexity

Suppose we want to minimize *f(x)* on a convex set *X*. Under very general conditions we can write *f* as the difference of two convex functions. It is sufficient to assume, for example, that *f* is twice continuously differentiable. It is necessary and sufficient that *f* is the indefinite integral of a function of locally bounded variation (Hartman,1959).

If *f=u-v,* with *u* and *v* both convex, then we use

$$v(x) \geq v(y) + Dv(y)(x-y)$$

to construct the convex majorization function

$$g(x,y) = u(x) - v(y) - Dv(y)(x-y).$$

The majorization method reduces optimization of an arbitrary function to solving a sequence of convex optimization problems. Of course matters simplify if *u(x)* can be chosen to be quadratic. For this majorization we find for the derivative of the algorithmic map

$$\left[D^2 u(x)\right]^{-1} D^2 v(x).$$

5. Using Taylor's Theorem

By Taylor's theorem

$$(5)\quad f(x) \leq f(y) + (x-y)'Df(y) + \frac{1}{2}\max_{0\leq\xi\leq 1}(x-y)'D^2 f(\xi x + (a-\xi)y)(x-y),$$

and the right hand side can be used as the majorization function. Of course this general approach can also be applied if we only use the linear term in the Taylor expansion, and also if we use third or higher order terms (De Leeuw, 2006). And by replacing max by min we can use it to construct minorization functions.

But let us continue with *quadratic majorization*. The majorization function in (5) is not necessarily simple, so we may want one that is easier to compute. Suppose there is a matrix *B* such that $D^2 f(x) \underset{\sim}{<} B,$, in the sense that *B-D²f(x)* is positive semi-definite for all *x*. Then clearly

$$g(x,y) = f(y) + (x-y)'Df(y) + \frac{1}{2}(x-y)'B(x-y)$$

is a majorization function for *f.*

By defining the current *target*

$$z = y - B^{-1}Df(y),$$

and by completing the square, we see that

$$g(x,y) = f(y) + \frac{1}{2}(x-z)'B(x-z) - \frac{1}{2}Df(y)'B^{-1}Df(y)$$

Thus step *k* of the majorization algorithm solves the least squares problem

$$\min_{x\in X}\left(x - z^{(k)}\right)'B\left(x - z^{(k)}\right).$$

We can choose the matrix *B* to be scalar, for instance by using an upper bound for the largest eigenvalue of *D²f(ξ)*. In that case computing the target simplifies, and all majorization subproblems are unweighted least squares problems.

In the case in which *X* is all of $\Re^n$ the quadratic majorization algorithm simply becomes

$$x^{(k+1)} = x^k - B^{-1}Df\left(x^{(k)}\right).$$

This algorithm will in general have a linear convergence rate $1-\lambda(x)$, where $\lambda(x)$ is the smallest eigenvalue of $B^{-1}D^2f(x)$ and x is the fixed point. A smaller B will give a more rapid convergence rate, but in general we cannot expect to see anything faster than linear convergence. If our bound B is really bad, then we may see very slow linear convergence.

6. Discussion

Majorization algorithms replace a complicated optimization problem by a sequence of simpler ones. In fact typically the subproblems are chosen in such a way that they are really simple to solve, and this is exactly what makes the algorithm attractive for really large problems such as the ones in tomography and microarray analysis. The quantities needed in an iterative step do not have a large footprint and can often be computed in parallel or from a stream of data. Convergence can be slow, and techniques to accelerate convergence may be necessary. But is often far more convenient to let a simple globally convergent ad-hoc algorithm run for hours than to try to fit huge and ill-conditioned matrices into memory in order to apply some suitably safe-guarded version of Newton's method.

References

1. J.De Leeuw. Quadratic and Cubic Majorization. Pre print series, UCLA Department of Statistics, 2006.
2. J. De Leeuw. Block Relaxation Methods in Statistics. In H.H. Bock, W.Lenski, and M.M. Richter, editors. *Information Systems and Data Analysis,* Berlin, 1994, Springer Verlag.
3. P. Hartman. On Functions Representable as a Difference of Two Convex Functions. *Pacific Journal of Mathematics*, 9:707-713,1959.
4. W.J.Heiser. Convergent Computing by Iterative Majorization: Theory and Applications in Multidimensional Data Analysis. In W.J. Krzanowski, editor, *Recent Advances in Descriptive Multivariate Analysis*, pages 157-189, Oxford: Clarendon Press, 1995.
5. K. Lange, D.R. Hunter, and I. Yang. Optimization Transfer Using Surrogate Objective Functions. *Journal of Computational and Graphical Statistics*, 9:1-20,2000.
6. J.M.Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables. Academic Press*, New York, N.Y., 1970.

Seagulls outside the February 2006 Program Chair Meeting in Alexandria, VA. (Photo courtesy of Juergen Symanzik)

"R has sounded the death knell for statistical computing research." Anonymous

Tools for Computing

RGTK2 - A GUI TOOLKIT FOR R

Michael Lawrence, Iowa State University
Duncan Temple Lang, University of California, Davis
lawremi@iastate.edu
duncan@wald.ucdavis.edu

1. Motivation

RGtk2 enables the R programmer to construct graphical user interfaces with GTK+, an open-source GUI toolkit. The R platform greatly benefits from access to GUI's that allow novice users, such as biologists, to capitalize on the analytical functionality of R, without the hindrance of the learning curve associated with a console-driven interface. For example, a graphical interface could guide a biologist through a microarray data analysis task using Bioconductor. GTK+ is a vast improvement over existing GUI toolkits for R, such as tcl/tk, because GTK+ is more advanced, does not sacrifice platform independence (available on Windows, Mac, and Linux) and, by virtue of its popularity in open-source applications, is capable of integrating interface functionality from a wide-range of other projects, including GGobi and Mozilla Firefox.

The original RGtk, based on GTK+ version 1.2, was developed by Duncan Temple Lang. About 4 years ago, GTK+ 1.2 was overhauled and renamed to GTK2. The fundamental GTK object system was abstracted into a separate library called GObject, part of GLib. Many widgets were added, removed, and heavily altered. GTK2 has more sophisticated widgets, prettier text, and a more elegant foundation than its predecessor. RGtk2, has caught up with the evolution of GTK+, with virtually complete support for the latest version of GTK2 (2.8.0) and its underlying libraries.

2. Features

RGtk2 offers many improvements over the original:

✓ A rewritten back-end based on the new GObject API.

✓ Automatic memory management, partly done in the original, is now completely implemented and automatic. The R programmer doesn't know that it's there.

✓ Beyond the new GTK2 widgets, there are complete bindings to many new GTK2-associated libraries, including Cairo, GdkPixbuf, ATK, Pango, and Libglade. Some of these are even useful independent of a GTK+ GUI. For example, the Cairo bindings allow the R programmer to draw vector graphics to PNG images. An extensive set of demos is available to introduce the user to the RGtk style.

✓ RGtkDataFrame object: A GTK data model that is backed by an R data frame, helpful when displaying large amounts of data in an interface. Its sorting ability, which delegates to R, is faster than that of the built-in GTK data model.

✓ A GTK implementation of Simon Urbanek's elegant cross-toolkit iWidgets API.

✓ Ability to embed Cairo-drawn R graphics. Cairo provides good quality,, albeit slow, anti-aliased graphics and text. It requires the cairoDevice R package, written by Michael Lawrence.

✓ Rd format documentation automatically derived from that of the bound libraries.

3. Design

The design goals of the project are two-fold. First, the bindings must be complete and consistent with the bound API. Whatever the C programmer can do, the R programmer should be able to do. Second, interaction with RGtk2 must be simple and familiar to the R programmer. Foreign C concepts such as memory management, return-by-reference parameters, and type casting must be hidden or adapted to their R equivalent. The user should be able to enjoy the benefits of GTK+ without knowing that it is implemented in a foreign language.

Like the original RGtk, the majority of the bindings are autogenerated from a description of the API's called *defs* files, while the rest are implemented manually. The *defs* files and their Python-based parser are provided by the pygtk project. The parser is invoked via RSPython, and the R and C code is generated in R from the parse result. Put simply, the bindings convert the input parameters from R to C, invoke the C function, and convert the result from C to R.

Thus, most of the work outside of autogeneration deals with converting different types. The primitive

types are straightforward, as are pointers to C structures (objects) constructed by the API. This was the extent of the original RGtk's type conversion. A more complicated problem that RGtk2 attempts to solve is the conversion of simple, transparent C structures that are normally initialized manually and therefore lack a constructor. Instead of defining a non-standard API function, the user is asked to define an instance of such a "transparent" type as an R list which is automatically converted to and from the C structure when passed between R and C. R closures are also wrapped on the fly to satisfy C "user function" parameters.

Other adaptations include allowing the user to leave off array length parameters (not necessary in R), and returning "out" or "return-by-reference" parameters along with the return value in an R list.

On Windows installation of gtk2 is simple, and on Mac OSX it is time-consuming but just as simple. On linux, the installation of recent versions of gtk2 can conflict with dependencies on earlier versions.

4. Impact and Future Work

In the several months since RGtk2's release, it has already been adopted by several projects. John Verzani has adapted his RGtk-based R GUI, Poor Man's GUI (PMG), to RGtk2. He has also expanded the iWidgets API to include many more widgets and has implemented the new functions with RGtk2. Written purely in R, PMG has great potential as an advanced, cross-platform GUI. It also promises to solve RGtk2's minor issues with the R event loop, since it allows the GTK main loop to run fulltime. The data mining tool Rattle, by Graham Williams, is based on Glade (an XML description of a GTK+ GUI) and was ported from Python by writing a single line of R code calling Libglade via RGtk2. This allowed Rattle to quickly capitalize on R's analytical functionality. Zhesi He is basing her bioinformatics data visualization project, Vitamin B, on RGtk2. Many other projects stand to greatly benefit from RGtk2, such as Elizabeth Whalen's RGtk-based iSPlot and iSNetwork packages in BioConductor, as well as any other R package that wishes to add a new dimension of usability through a GUI, without suffering the inconveniences and limitations of the tcltk package. Work has also begun with Daniel Adler to replace the polymorphic RGL backend with one based on GtkGLExt, which will be simpler and also allow integration of RGL visuals with RGtk2 interfaces.

Future plans for RGtk2 include improvement of code autogeneration (less manual work), allowing the R implementation of custom GObjects, and keeping pace with new GTK releases, which may soon support full introspection, allowing us to bind arbitrary GObject libraries on the fly.

References

1. RGtk2 website: http://www.ggobi.org/rgtk2
2. Original RGtk: http://www.omegahat.org/RGtk
3. GTK+: http://www.gtk.org/
4. Poor Man's GUI: http://www.math.csi.cuny.edu/pmg
5. Rattle: http://www.togaware.com/datamining/rattle.html
6. Vitamin B: http://www-users.york.ac.uk/~zh107/phd.html
7. BioConductor: http://www.bioconductor.org/
8. RGL: http://rgl.neoscientists.org/
9. GtkGLExt: http://gtkglext.sourceforge.net/
10. PyGTK: http://www.pygtk.org/
11. exploRase: http://croc.vrac.iastate.edu/explorase/
12. Temple Lang, Duncan. RSPython: http://www.omegahat.org/RSPython.

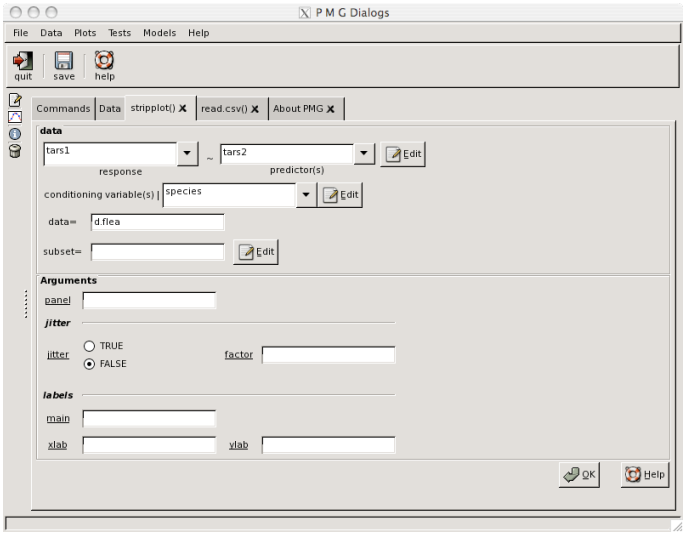


Figure 1. An example of RGtk2 usage in the pmg implementation by John Verzani.

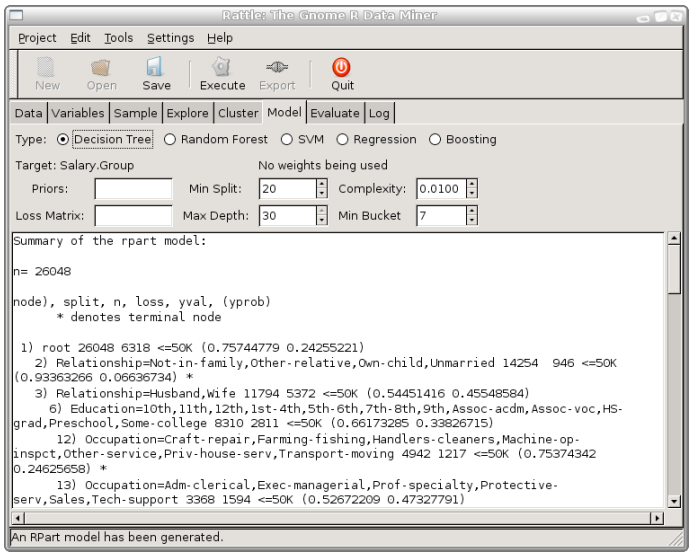


Figure 2. An example of RGtk2 usage in the Rattle implementation.

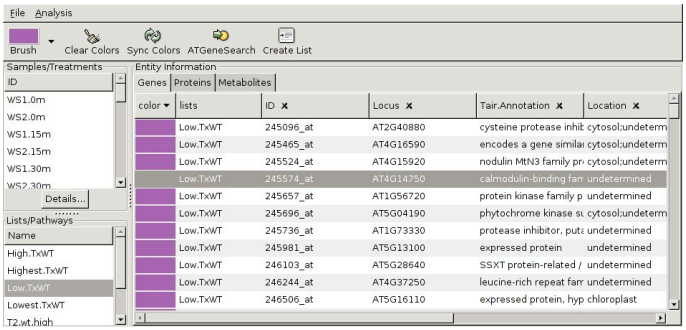


Figure 3. An example of RGtk2 usage in the exploRase implementation.

"R serves as a mission of service and sacrifice to the statistical community. Rather than wasting computing talent, it serves to accelerate statistical research." Anonymous

STATISTICAL COMPUTING AT JSM 2006

by Michael Trosset, 2006 Program Chair
Statistical Computing Section

The Section on Statistical Computing will be the primary sponsor of eight invited, four topic contributed, and fifteen contributed sessions at JSM 2006, to be held August 6--10, in Seattle, at the Washington State Convention & Trade Center. Here is the schedule of Invited and Topic Contributed sessions:

- Sunday, Aug 6, 2:00-3:50pm: Density-Based Clustering (organized by David Scott)
- Sunday, Aug 6, 4:00-5:50pm: Issues with Open Source Statistical Software in Industry (organized bu Nicholas John Lewin Koh)
- Monday, Aug 7, 8:30-10:20am: Statistical and Computational Issues in Climate Research (organized by Don Percival)
- Monday, Aug 7, 10:30am-12:20pm: Machine Learning and Beyond: A Session in Memory of Leo Breiman (organized by Liza Levina)
- Monday, Aug 7, 2:00-3:50pm: Genome-wide Association Studies (organized by Charles Kooperberg)
- Tuesday, Aug 8, 10:30am-12:20pm Personal Networks: Applications Using Data on Social Relationships (organized by Chris Volinsky)
- Tuesday, Aug 8, 10:30am-12:20pm: Student Paper Award Winners (organized by Jose Pinheiro)
- Tuesday, Aug 8, 2:00-3:50pm: Computational Challenges of Massive Data Sets and Sources (organized by Karen Kafadar)
- Wednesday, Aug 9, 8:30-10:20am: New Directions in Statistical Machine Learning (organized by Yufeng Liu)
- Wednesday, Aug 9, 10:30am-12:20pm: Multidimensional Scaling and Manifold Learning (organized by Michael Trosset)
- Wednesday, Aug 9, 2:00-3:50pm: Least Angle Regression (organized by Tim Hesterberg)
- Thursday, Aug 10, 10:30am-12:20pm: Monte Carlo Methods for Computationally Intensive Problems (organized by Yuguo Chen)

A preliminary program for the entire meeting is posted at

http://www.amstat.org/meetings/jsm/2006/index.cfm

Tools for Teaching

SOCR: STATISTICS ONLINE COMPUTATIONAL RESOURCE: SOCR.UCLA.EDU

Ivo D. Dinov
Department of Statistics & Center for Computational Biology, University of California Los Angeles
dinov@stat.ucla.edu
http://www.socr.ucla.edu/

1. Background

Statistical computing commonly involves data design, acquisition and integration, model development, analysis, visualization and results interpretation [2-7]. In the past several years, a number of groups across the globe have introduced various interactive Internet-based statistical computational resources and educational tools. Among these are: Seeing Statistics, SurfStat, Goose Statistics Environment, StatSoft, HyperStat, Statistics at Square One, WassarStats, ResamplingStats, WebStat, RJava, CUWU Stats, PSOL, StatLab, Virtual Labs in Probability & Statistics, JavaStat, Vistac, JSci, CyberStat, JFreeChart, etc. Some of these have general theoretical treatments, whereas others have very specialized flavors. There are varying amounts of contextual information, degrees of computational capability, interactivity and portability among these tools, and data- or tool-interaction is sometimes limited.

2. The SOCR Resource

The computational tools in the SOCR resource include a collection of Java applets, user interfaces and demonstrations divided in seven major categories of resources: interactive distribution modeler, virtual experiments, statistical analyses, computer generated games, data modeler, chart and graphing tool and a collection of additional tools. Science magazine has published a brief review of our SOCR probability and statistics resources [8].

SOCR Distributions: The SOCR suite of distribution modeling aids includes tools for sampling/resampling, hypothesis testing, statistical inference, model-fitting and critical value estimation.

This SOCR Distribution component (http://www.socr.ucla.edu/htmls/SOCR_Distributions.html) allows interactive manipulation of over 35 different families of distributions. Even though the benefit of this tool is limited in some data analysis situations (by the accuracy of the interactive hand-motion, mouse-precision and screen-resolution) we have received a strong and clearly positive feedback from users and experts on the intuitive nature, flexible design and the pedagogical potential of this interactive distribution modeling toolbox. Figure 1 illustrates one example with the (general) Beta (A=4.0, B=11.0, α=6.0, β=2.0) distribution. This applet allows interactive manipulations of the distribution parameters and direct calculations of areas of interest by manually clicking and dragging the mouse on the canvas. It reports some basic statistics in the text area on the bottom.

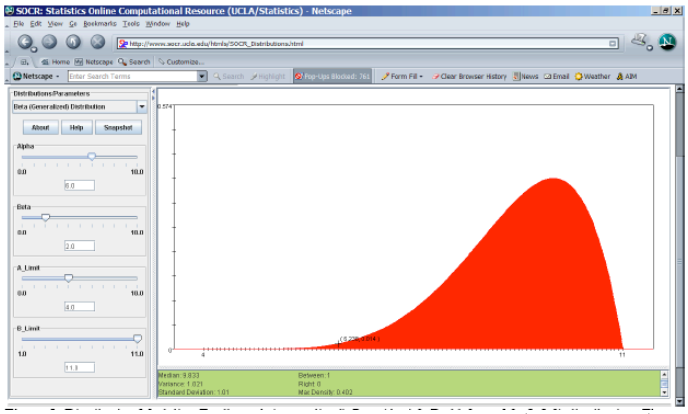


Figure 1: *Distribution Modeling Toolbox: A (generalized) Beta (A=4,B=11.0, α=6.0,β=2.0) distribution. the tool allows interactive update of the four parameters and directly computes areas of interest.*

SOCR Games: Many data processing and analysis protocols rely on frequency-based transformations. In a classical setting, data are often times observed in, transformed to, or analyzed in the Fourier space [9]. We have provided a novel design of exploiting multi-dimensional characteristics of signals in the spatial and frequency domains.

This tool is part of the SOCR games (http://www.socr.ucla.edu/htmls/SOCR_Games.html) and allows users to first generate 1D signals (e.g., audio) and then investigate the effects of signal changes onto their Fourier and Wavelet space representations [10].

Conversely, determining the effects of various frequency, location and intensity-magnitude effects on the special characteristics of the data can also be explored. Figure 2 depicts one such example where Fourier space representation of signals is visually demonstrated. The Fourier transformation of the signal is instantly computed and displayed, which allows the users to monitor the effects of altering the magnitudes or phases of the spectral coefficients of the original signal. A mouse-over event in the bottom Fourier-space panel generates the exact magnitude effect of the signal in the spatial domain. All these manipulations are controlled intuitively in real-time by the user using a mouse.

SOCR Experiments: We have also begun the high-level design and modeling of Java applets that illustrate more advanced statistical methods and techniques. These include general linear modeling, power analysis, expectation maximization, likelihood ratio inference, mixture modeling, stochastic integration, Brownian

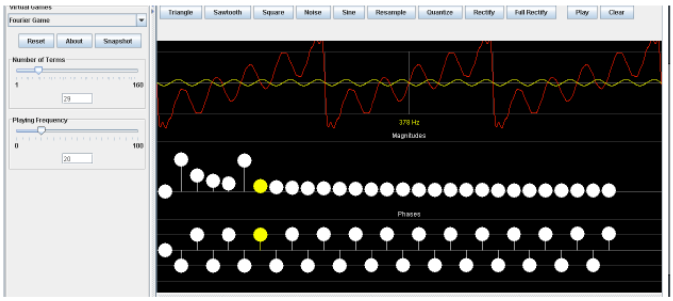


Figure 2: *A Virtual Fourier Game. It illustrates the dynamic, user-controlled interplay between the signal-intensities and the spatial, frequency, phases and magnitudes of the Fourier coefficients.*

motion, Markov Chain Monte Carlo methods, etc. One example that shows both interactive distribution-mixture-modeling and generalized-expectation-maximization was implemented in the setting of 2D point clustering and classification. Figure 3 demonstrates one example of the SOCR Experiments (http://www.socr.ucla.edu/htmls/SOCR_Experiments.html). It fits a 3-term Gaussian mixture model with random isotropic starting kernels in 3D, using expectation maximization for parameter estimation (final kernel positions, shapes and point classification). This methodology is applied to classify brain tissue types dynamically in the mouse brain. The LONI Viz viewer [11], a stand-alone application independent of SOCR, integrates this SOCR functionality by calling the appropriate computational methods inside the SOCR EM modeling engine [12].

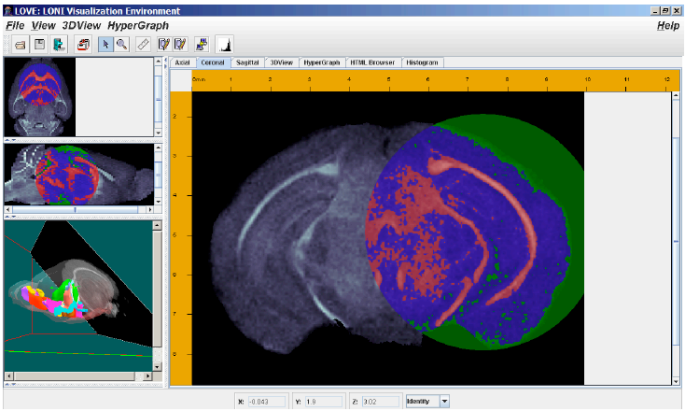


Figure 3: *SOCR Expectation Maximization (RM) and Mixture Modeling: Dynamic brain tissue segmentation within LONI Viz (see text) which is accomplished using the SOCR EM utilities. Instantaneous classification of volumetric brain data into white matter, gray matter and cerebrospinal fluid is obtained for the MRI image of the mouse brain.*

SOCR Modeler: Another component of SOCR deals with the issue of model fitting. We have designed the SOCR-Modeler-framework (http://www.socr.ucla.edu/htmls/SOCR_Modeler.html), with a flexible data input, which fits a desired distribution model to the data, see Figure 4. This approach does not necessarily produce an optimal model. However, one can develop various strategies for analytically optimal computer-based model estimation strategies in some of these situations.

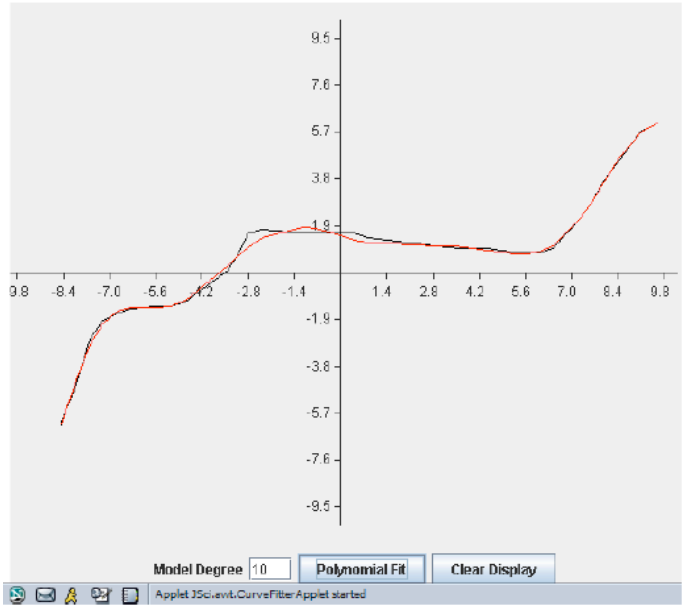


Figure 4: *SOCR Modeler: A polynomial or a distribution model may be fit to manually drawn data. The analytical model and the quality of the fit are then reported, based on the user selected polynomial degree.*

SOCR analyses (http://www.socr.ucla.edu/htmls/SOCR_Analyses.html) consist of a cluster of tools for real data analysis. Figure 5 depicts one of the several analyses schemes that we have developed. In this example a random sample, from our distribution utility, is generated (dependent values) and a 2-way analysis of variance is performed to identify potential main and simple effects for the two predictors (factors A and B).

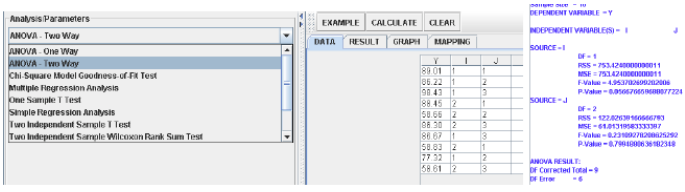


Figure 5: *SOCR Analyses: 2-Way Analysis of Variance (ANOVA) example on randomly generated data.*

SOCR Charts: The sixth component of SOCR is a collection of chart and graphing tools based on JFreeChart.

SOCR Charts (http://www.socr.ucla.edu/htmls/SOCR_Charts.html) include a large number of commonly user plots and graphics and provide some data summary statistics. Finally, SOCR includes a collection of additional resources developed by other groups (http://www.socr.ucla.edu/Applets.dir/OnlineResources.html).

3. SOCR Resource Infrastructure

We currently use a 2 GHz dual-processor Mac OS X server to support the SOCR resource at http://www.socr.ucla.edu. Over 100GB of hard disk space and 1 Giga-bit-per-second Internet connection make the SOCR resource robust and responsive. Still, we have a number of developments and improvements to make to complete the SOCR resource according to the specified design. Any remote client (user) having Internet connection and a Java-enabled web-browser can access the SOCR functionality from any place in the world and at any time of the day. However, there are still great differences in computer hardware, java-virtual machines, browser versions & manufactures, user settings and firewall/privacy/security preferences that are allowed by modern operating systems, web-browsers and Internet providers. This variation of settings often times causes problems for some users to access the SOCR resource. We have tested and verified that the SOCR tools are accessible via Java 1.2+ on Mac OS X, IRIX, Solaris and Windows platforms under Netscape 7.2 +, FireFox 1.5 and InternetExplorer 6.0+. In addition, SOCR is UCLA-signed (Thawte [13]) as a secure applet which may cause problems for some users within restrictive firewall networks. Signing was required for interactive data input and output and mouse buffer (cut-and-paste) functionality.

The SOCR end-user documentation iscurrently being designed, but many of the SOCR tools include Help and About buttons with the corresponding functions. The SOCR developer documentation is already available at the SOCR documentation web-page as Java-class hierarchy and as UML diagrams. Currently SOCR has a very small kernel and all available features and applications are of plug-in type. This allows

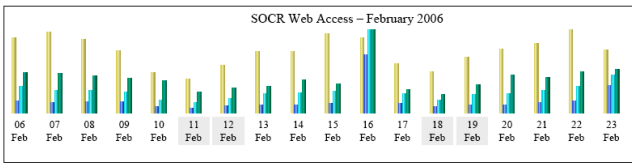
direct SOCR expansion by outside groups and investigators. The SOCR resource is open-source, we share source and binary code and welcome any contributions from the community.

A SOCR online discussion forum is also deployed for people to collaborate, share their ideas and possibly form special interest groups, e.g. course- or topic-specific interests. The discussion forum is moderated from time to time for content and suggestions. This helps the online community share their ideas and provide valuable feedback for future developments and improvements to the SOCR environment. Specific meta-searches (based on date, keyword and subject) of the SOCR materials, tools, resources, repositories and forums are provided via a site-specific Google search engine.

The SOCR resource has attracted over 60,000 visitors since 2002. This large number of users translated into over 30,000 actual active users. These are statistics of users, not hits, and include a single event counting per user per day, no matter how many resources or tools were utilized by the visitor. These numbers also exclude the visits to our educational materials, notes and tutorial provided as additional resources and linked to from SOCR. Typical users consisted of UCLA undergraduate students, local researchers and outside students and investigators. We have summary statistics of the hourly, daily and monthly SOCR usage statistics, as well as browser, operating system, country or origin and type of access to the resources. These are available online at the SOCR Acknowledgement page and an example is illustrated on Table 1.

Table 1: Overall and daily statistics of the SOCR Resources traffic for the month of February 2006.

Reported period	Month Feb 2006				
First visit	01 Feb 2006 - 00:19				
Last visit	28 Feb 2006 - 23:58				
	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Viewed traffic *	2890	3617 (1.25 visits/visitor)	17532 (4.84 pages/visit)	32051 (8.86 hits/visit)	2.13 GB (616.48 KB/visit)



A number of outside academic, non-profit and industrial resources were also utilized in the SOCR development. We have used ideas, design, tools and models from Elementary Statistics Java Applets [14], Statlets [15], Rice Virtual Labs in Statistics [16], StatCrunch [17], Statiscope [18], PsychStat [19], BusinessStat [20], Probability by Surprise [21], Web Interface for Statistics Education [22], CUWU Stats [23], PSOL [24], StatLab [25], Virtual Labs in Probability & Statistics [26], JavaStat [27], Vestac [28], JSci [29], CyberStats [30], and many other groups, organizations, student projects, research, clinical and teaching resources.

Acknowledgements

Many people have contributed to the SOCR development efforts in one form or another over the years. These include, among others, Petros Efstathopoulos, Juana Sanchez, Nicolas Christou, Robert Gould, Guogang Hu, Jianming Hu, Jason Landerman, Fotios Konstantinidios, Hui Wang, Donald Ylvisaker, Dushyanth Krishnamurthy, Jeff Ma, Annie Che, Jenny Cui and Arthur Toga.

In addition, the SOCR project was supported in part by the following State of California and Federal grants UCLA OID IIP (IIP0318), NIH/NCRR (P41 RR13642), NCBC (U54 RR021813) and NSF CCLI-EMD (044299).

References

1. Dinov, I., *Online Probability and Statistics Computational Tools:* *http://www.socr.ucla.edu/htmls/SOCR_Acknowledgments.html*. 2006.
2. Whitley, E. and J. Ball, *Statistics review 2: samples and populations.* Crit Care, 2002. **6**(2): p. 143-8.
3. Whitley, E. and J. Ball, *Statistics review 3: hypothesis testing and P values.* Crit Care, 2002. **6**(3): p. 222-5.
4. Whitley, E. and J. Ball, *Statistics review 5: Comparison of means.* Crit Care, 2002. **6**(5): p. 424-8.
5. Whitley, E. and J. Ball, *Statistics review 6: Nonparametric methods.* Crit Care, 2002. **6**(6): p. 509-13.
6. Whitley, E. and J. Ball, *Statistics review 1: presenting and summarising data.* Crit Care, 2002. **6**(1): p. 66-71.
7. Dianne, C., Maitra, R., *Statistical Computing.* Newsletter of the ASA Statistical Computing & Statistical Graphics, 2005. **16**(1): p. 1-3.
8. Leslie, M., *Statistics Starter Kit.* Science, 2003. **302**(5): p. 1635.

9. Xu, Z.H. and A.K. Chan, *Encoding with frames in MRI and analysis of the signal-to-noise ratio.* Ieee Transactions on Medical Imaging, 2002. **21**(4): p. 332-342.
10. Dinov, I.D., et al., *Quantitative comparison and analysis of brain image registration using frequency-adaptive wavelet shrinkage.* IEEE Trans Inf Technol Biomed, 2002. **6**(1): p. 73-85.
11. Dinov ID, V.D., Shin BC, Konstantinidis F, Hu G, MacKenzie-Graham A, Lee EF, Shattuck DW, Ma J, Schwartz C and Toga AW., *LONI Visualization Environment.* Journal of Digital Imaging, 2006. **in press**.
12. Krishnamurthi, D., *Development of Statistical Online Computational Resources and Teaching Tools*, in *Statistics*. 2005, UCLA: Los Angeles. p. 47.
13. Thawte, *http://www.thawte.com*.
14. ElementaryStats, *http://intrepid.mcs.kent.edu/%7Eblewis/stat/*.
15. Statlets, *http://www.statlets.com/*.
16. RVLS, *http://www.ruf.rice.edu/%7Elane/rvls.html*.
17. StatCrunch, *http://www.statcrunch.com/*.
18. Statiscope, *http://www.df.lth.se/~mikaelb/statiscope/statiscope.shtml*.
19. PsychStat, *http://www.psychstat.smsu.edu/*.
20. BusinessStat, *http://home.ubalt.edu/ntsbarsh/zero/scientificCal.htm*.
21. ProbBySurprise, *http://www-stat.stanford.edu/%7Esusan/surprise/index.html*.
22. WebStatEducation, *http://wise.cgu.edu/*.
23. CUWUStats, *http://www.stat.uiuc.edu/courses/stat100//cuwu/*.
24. PSOL, *http://www.math.uah.edu/stat/objects/index.xml*.
25. StatLab, *http://statlab.fon.bg.ac.yu/eng/eng/apletieng/resources/resources4.html*.
26. VirtualLabs, *http://www.math.uah.edu/stat/*.
27. JavaStat, *http://www.umd.umich.edu/casl/socsci/econ/StudyAids/JavaStat/applet.htm*.
28. Vestac, *http://www.kuleuven.ac.be/ucs/java/*.
29. JSci, *http://jsci.sourceforge.net/*.
30. CyberStats, *http://statistics.cyberk.com/splash/*.

STATISTICAL GRAPHICS AT JSM 2006

by Juergen Symanzik, 2006 Program Chair
Statistical Graphics

Stat Graphics was very successful in the lottery (sorry, the Section Competition) where three of our suggested invited sessions had to compete against suggested invited sessions from other sections. All three of our suggested sessions for the competition were successful - plus two guaranteed invited sessions, giving us the maximum possible of five invited sessions at the JSM this year, one each day. In chronological order, these are:

- **Sun 08/06/2006**, 4:00 PM to 5:50 PM:``Statistical Graphics: from Playfair to Bertin and Beyond"
- **Mo 08/07/2006**, 10:30 AM to 12:20 PM:``Statistical Graphics: Applications in Drug Discovery and Clinical Development"
- **Tu 08/08/2006**, 8:30 AM to 10:20 AM:``Network Visualization"
- **We 08/09/2006**, 10:30 AM to 12:20 PM:``Human Perception and Statistical Graphics"
- **Th 08/10/2006**, 8:30 AM to 10:20 AM:``Graphical Tools for Spatial Econometrics"

Another highlight is the data expo, held for the first time in many years. If you did not hear about this expo before, take a look at

http://www.amstat-online.org/sections/graphics/dataexpo/2006.php.

We have 14 poster entries presented in a topic contributed data expo poster session on **Mo 08/07**/2006, 10:30 AM to 12:20 PM. Take a close look at these posters, judge them, and decide for yourself who will be awarded first, second, and third prizes. The first prize consists of $1000 cash plus a set of NASA books, the second prize is $500 and a set of books, and the third prize is $200 plus a set of NASA posters. If you want to know who the winners will be, the prizes will be awarded at the joint ``Statistical Graphics and Statistical Computing Sections Mixer" on **Mo 08/07/2006**, 7:30 PM to 10:00 PM. In addition, there will be lots of free food, drinks, and door prizes to be won. If you want to have fun, meet with colleagues from Stat Graphics and Stat Computing you usually see only once a year (here!), or just enjoy the food, the Mixer is the place to be.

Continues on Page 21.....

From the Field

WHAT DO STATISTICAL COMPUTING AND GRAPHICS FOLKS DO?

RESEARCH DEPARTMENT AT INSIGHTFUL

Tim Hesterberg

In this article I'll describe the work environment in the Research Department at Insightful, creator of S-PLUS software. The intended reader is someone who enjoys computational statistics and research, and possibly consulting and travel, and would like to read about opportunities other than academia.

I joined the department in 1996, to work on an NIH-funded projec for developing resampling software, after eight years in academia.

My main research focus was on computationally-efficient methods for bootstrap calculations, and this was an opportunity to turn that research into software that is widely usable. With the aid of additional funding from NSF, that project resulted in the S+Resample library, as well as book chapters for use in teaching statistics; see www.insightful.com/Hesterberg/bootstrap. I've since been involved in a number other projects offering a great deal of variety, including sequential designs for clinical trials, missing data using multiple imputations, stable distributions, simulation-based econometric software, visualization for multiple-bandwidth smoothing, functional data, long-memory data, image analysis, and seismic deformation estimation.

The department is primarily externally funded, with grants from NIH, NSF, NASA, and defense agencies. We primarily compete for Small Business Innovation Research (SBIR) grants, which require research in areas with commercial potential. We typically work with academics consultants who have developed basic methodology in a new area and perhaps software, and expand the methodology and software, fill in the gaps, add help files and manuals, and do a lot of testing. (The expectations for software quality here are much higher than I was used to as an academic)

These SBIR grants are probably easier to get than typical academic research grants, because Congress likes to encourage small businesses. We've had a good success rate in applications, and the department is looking to grow; we currently have openings in machine learning and mixed effects models, with other possibilities coming up (contact me if you're interested).

The work involves a combination of writing grants, algorithmic and other research, software engineering, creating case studies, and writing documentation, technical reports and articles, and reports to the funding agencies.

The work environment is normally sane, though long hours are common when writing grants or when finishing projects by deadlines. Publications are encouraged, as part of a project or on your own time, see e.g. www.insightful.com/Hesterberg/articles.

I've found support for attending conferences easier here than as an academic.

There are also opportunities for consulting work. I've done consulting related to research projects in clinical trials and resampling, and some other consulting, including a month last summer on a risk management project in Zurich (a real hardship post!).

There are also opportunities for teaching short courses. I've taught courses related to my research, on bootstrapping and permutation tests. This provides an opportunity work toward two of my dreams -- to change how statistics is taught at the introductory level (more simulation, bootstrapping, and visualization, less cookbook formulas), and how it is practiced (less blind reliance on the central limit theorem, and more checking whether sampling distributions are actually normal and unbiased by using bootstrapping). This also involves travel, to such exotic locations as Rochester MN and Little Rock. Oh yes, also Albuquerque, San Francisco, Boston, Chicago, L.A., Washington D.C., Minneapolis, Cincinnati, Toronto, London, Manchester, Basingstoke, Zurich, Basel, Montpellier, Frankfurt and Paris. It ain't bad, if you enjoy travel.

Finally, there's money. I earn substantially more here than as an academic. While I don't care a lot about possessions (and you probably don't either, if you're in or considering the Ph.D./academia route), it is nice to be able to give large amounts to things I care about, like schools, the Sierra Club, and the congressional campaign of college friend I respect.

Publications

THE JOURNAL OF TECHNOLOGICAL INNOVATIONS IN STATISTICS EDUCATION

Robert Gould and Nathan Yau
UCLA Department of Statistics

The UCLA Department of Statistics' Center for Teaching Statistics is launching a new journal devoted to the intersection of statistics education and technology. *Technology Innovations in Statistics Education* (TISE), will meet what we feel is a critical need for a scholarly journal devoted to how to teach technology and how to teach with technology.

Teachers of statistics are faced with decisions at all levels about how to best use technology. We use technology in our presentations (overheads, software demonstrations, simulations), to structure our courses, (course management systems, distance learning tools), and as part of assignments to our students (simulations, data analyses). Sometimes the technology is secondary; we use an applet to teach the central limit theorem not because we wish to teach applets, but because we feel it's the best way to build students' intuition about the central limit theorem. Other times, the technology is primary; at UCLA we have computer labs that teach STATA and R because we want our students to become adept at using statistical software. At some advanced levels, we teach the students to become designers and creators of technology so that they can access and analyze new and complex data structures.

If we think of a statistics education as an education in data science, then technology is not just an aid to learning, but is a subject that must itself be learned. Data scientists must be able to use technology to reason and think with data. Data scientists must be able to design technology to access data and teach others about data. Educators must know how to best teach these students to use and fashion technology.

Technology Innovations in Statistics Education (TISE) will be an on-line, peer-reviewed journal for scholarly papers that research issues of technology in statistics education. The journal will be the first of its kind. While there are a large number of journals that explicitly address issues in educational technology (Educational Media International, Teaching with Technology Today, Journal of Management Information Systems, Journal of Educational Media, Journal of Computer Assisted Learning, Journal of Online Learning and Teaching to name a few), there are only a small number of journals devoted to statistics education: the Journal of Statistics Education (JSE) (published by the ASA), the Statistics Education and Research Journal (SERJ) (published by IASE), and Teaching Statistics (published by the Royal Statistical Society). There are slightly more if we include other academic journals that sometimes publish papers on statistics education or teaching-related matters. The first group of journals focuses on technology, but rarely publish papers related to statistics education. The statistics education journals do not focus on technology and yet publish a fair number of articles that address technology issues, but with significant gaps.

The JSE is probably the best known of the statistics education journals, certainly in the United States. The JSE's mission is broad and includes "the improvement of statistics education at all levels." JSE publishes three issues per year and is currently in its 14th year. According to its web page (http://www.amstat.org/publications/jse/ accessed May 23, 2006) JSE receives about 100 papers per year and has an acceptance rate of about 20%. As this "word cloud" shows, technological words (simulation, http, software, computer) are fairly frequent throughout the first fourteen volumes. (The placement of the words is arbitrary. The size of the words is proportional to their frequency in the set of all JSE articles published up to Volume 14, Number 1 (March 2006)).

A textual analysis of the articles in JSE revealed four content "clusters", which we broadly define as "statistics methods and techniques", "probability and mathematical statistics", "educational" and "datasets and stories". Papers in the "statistical methods and techniques" section examined the teaching of particular methods, say ANOVA. An example of a paper in this cluster that discusses technology is "A Visualization Tool for One- and Two-way Analysis of Variance", (Rachel Sturm-Beiss, v13n1). Sturm-Beiss demonstrates the use of a java applet to help students relate the ANOVA table to features of the model. Papers in the "probability and mathematical statistics" section are more keyed towards fundamental concepts. For example, the paper "Using Simulation to Teach Distributions" (David P. Doane, v12n1) provides spreadsheet-based modules that make some abstract

concepts more concrete. This cluster contains the greatest concentration of technology-related papers. The educational cluster contains articles about assessment, curriculum, teaching methods, etc. There are very few technology-related articles, but "Using Computer Simulation Methods to Teach Statistics: A Review of the Literature" (Jamie D. Mills, v10n1) is an example. The fourth cluster is a designated category of JSE. The "Datasets and Stories" section of the JSE consists of actual data along with documentation that illustrates how the data are used in the classroom. Technology is sometimes mentioned in these papers, but usually simply by mentioning the software used to analyze the data.

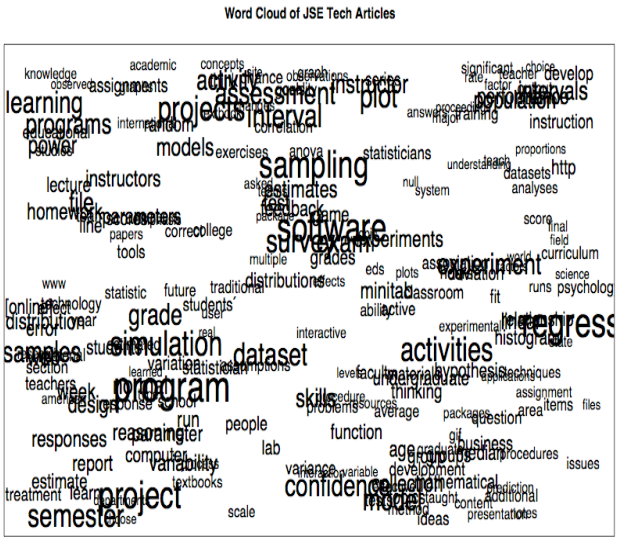


Figure 1. *A word cloud of the 255 most frequent words (excluding "noise" terms) in JSE. Position is arbitrary, but size is proportional to frequency.*

What's missing? Generally, the technology-related articles in JSE are directed towards helping educators understand how to use technology to help students understand fundamental concepts. In fact, we might argue that this is what the bulk of the literature on statistics education and technology is about. While important, this emphasis on using technology as a learning tool takes us only partway towards educating data scientists. We would like to see more papers on how to teach students a statistical software package, when to teach it, and what features of a package are beneficial and detrimental for learning to understand data. While it is important to learn how to teach with a piece of technology, say a certain applet, we would like to see papers on how and why that applet's design meets its educational objectives. We would like to see discussions of how to incorporate teaching technology into the curriculum, and how the curriculum is affected by changes in technology.

TISE will publish papers that help us use technology to help students better understand statistical concepts, that demonstrate how technology can help students gain insight from data, and that help us teach students to design and shape technology for future needs.

The first issue of TISE will appear in May 2006. You can "subscribe" (no charge) and be notified when the first issue appears (and other milestones) at http://tise.stat.ucla.edu. The editorial board is Robert Gould (founding editor), Mahtash Esfandiari, Christine Franklin, Joan Garfield, Brian Jersky, Joy Jordan, Cliff Konold, Deborah Nolan, Dennis Pearl, Roxy Peck, Duncan Temple Lang, Roger Woodard, Katie Makar, and Nathan Yau (assistant editor).

JOURNAL OF STATISTICAL SOFTWARE

Di Cook
Statistics, Iowa State University

The Journal of Statistical Software (www.jstatsoft.org/) hosted by UCLA has become a fully independent journal o the American Statistical Association. Abstracts will still appear in the Journal of Computational and Graphical Statistics. The publication quality has improved dramatically in recent years. If you are an academic encourage your institution to recognize publications in this journal as important for promotion. The nod from ASA should be sufficient to make the argument, and if there is still doubt emphasize the star status of associate editorial board!

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS

July 1, 2006 sees several changes for JCGS. Professor David Van Dyk, University of California-Irvine. take over the editor's position from Professor Luke Tierney. JCGS also moves to the same electronic review system that is used by other ASA journals.

News

RESULTS OF THE ANNUAL COMPETITIONS

Jose Pinheiro,
Awards Officer, 2006
Statistical Computing Section

The Statistical Computing Section of ASA sponsors three annual competitions aimed at promoting the development and dissemination of novel statistical computing methods and tools: the Student Paper competition (jointly with the Statistics Graphics Section), the John M. Chambers Award, and the Best Contributed Paper competition. Winners of all three awards are selected prior to the Joint Statistical Meetings (JSM), being officially announced at the Monday night business meeting of the Statistical Computing and Statistical Graphics Sections at JSM.

The Student Paper competition is open to all who are registered as a student (undergraduate or graduate) on or after September 1st of the previous year when the results are announced. Details on submission requirements are provided in the competition's announcement, which goes out in mid to late September, and at the Statistical Computing website at http://www.statcomputing.org. The four winners of the Student Paper competition are selected by a pannel of judges formed by the Council of Sections Representatives (COS-REPs) of the Statistical Computing and Statistical Graphics Sections, who work hard to get the results announced by the last week of January. As part of the award, the winners receive a plaque, have their JSM registration covered by the sponsoring sections and are reimbursed up to US$ 1,000 for their travel and housing expenses to attend the meetings. The winning papers are presented at a special Topics Contributed session at JSM, which typically takes place on Tuesday. The winners of the 2006 Student Paper competition, presented in alphabetical order, are:

- **Youjuan Li**, University of Michigan-Ann Arbor (advisor: Ji Zhu) "Efficient Computation and Variable Selection for the L1-norm Quantile Regression"

- **Fan Lu**, University of Wisconsin-Madison (advisor: Grace Wahba) "Kernel Regularization and Dimension Reduction"

- **Rebecca Nugent**, University of Washington-Seattle (advisor: Werner Stuetzle) "Clustering with Confidence"

- **Philip Reiss**, Columbia University (advisor: Todd Ogden) "An Algorithm for Regression of Scalars on Images"

The John M. Chambers Award is endowed by Dr. Chambers generous donation of the prestigious Software System Award of the Association for Computing Machinery presented to him in 1998 for the design and development of the S language. The competition is open to small teams of developers (which must include at least one student or recent graduate) that have designed and implemented a piece of statistical software, with the winner being selected by a panel of three judges, indicated by the section's awards officer. Further details on the requirements for submission and eligibility criteria are provided in the competition's announcement, which is distributed in early October, and at the Statistics Computing website (see above). The prize includes a plaque, a cash award of US$ 1,000, plus a US$ 1,000 allowance for travel and hotel expenses to attend JSM (with registration fee covered by the section.) The winner of the 2006 John M Chambers Award is:

- **Hadley Wickham**, Iowa State University (advisors: Di Cook and Heike Hofmann) "ggplot and reshape: Practical Tools for Organizing, Summarizing, and Displaying Data" (http://had.co.nz/jca2006)

Last, but not least, the Best Contributed Paper award is determined on the basis of the evaluations filled out by the attendees of the Contributed and Topics Contributed sessions of JSM which have the Statistical Computing Section as first sponsor. All presenters in those sessions are automatically entered in the competition. The prize includes a US$ 100 cash award and a plaque. The winner of the 2005 Best Contributed Paper Award is

- **Heather Turner,** Research Fellow at the Department of Statistics, University of Warwick, UK, (jointly with David Firth, from the same department) for the paper "Multiplicative Interaction Models in R" in the session "Algorithms and Software".

Continues from Page 15....

STATISTICAL GRAPHICS AT JSM 2006

Juergen Symanzik, 2006 Program Chair
Statistical Graphics

In addition to these sessions, we also have one topic contributed (TC) and two regular contributed sessions, as follows:

- **We 08/09/2006,** 2:00 PM to 3:50 PM``Visualization of Large Datasets" (TC)
- **Mo 08/07/2006**, 2:00 PM to 3:50 PM:``Advances in Graphical Methods"
- **Th 08/10/2006,** 10:30 AM to 12:20 PM:``Applications of Statistical Graphics"

Of course, we also have the ``Student Paper Award Winners" (joint with Statistical Computing) on **Tu 08/08/2006**, 10:30 AM to 12:20 PM. Is there a Stat Graphics paper among the four best student papers this year, out of about 20 papers submitted to the ``Student Paper Competition"? Check yourself! If you (or your students) are not among the winners this year, check out *http://www.statcomputing.org/awards/student/announcement.html* and get ready for the next round. Entries are usually due in mid to late December (2006).

Of course, we have the usual roundtables - with coffee roundtables offered for the first time this year:

- **Mo 08/07/2006**, 7:00 AM to 8:15 AM:``Graphics for Data Mining"
- **Tu 08/08/2006**, 12:30 PM to 1:50 PM:``Are Graphics/Interactive Graphics Useful for Getting Your Work Done?"
- **We 08/09/2006**, 12:30 PM to 1:50 PM:``Biostatistical Graphics: Large, Weak Datasets"

Moreover, we have a Continuing Education (CE) course titled ``Creating More Effective Graphs" on **Mo 08/07/2006**, 1:00 PM to 5:00 PM, joint with Statistical Education.

What - this is still not enough for your JSM program? OK, Stat Graphics co-sponsors about 15 sessions from other sections. Those are sessions that complement our sessions (e.g., several other sessions on networks) or entire sessions or individual talks that are related to graphics. For example, the first talk in the Statistical Education session titled ``Teaching Statistics to Specific Audiences" on **Tu 08/08/2006,** 2:00 PM to 3:50 PM, is titled ``Teaching Effective Graph and Table Construction Needs More Attention in Statistical Education." I think we can all agree! By the way, this is not suggested by instructor X at college Y, but by someone from the pharmaceutical industry!

So, take a careful look at the online program or the printed program later at the JSM and decide which of the sessions co-sponsored by Stat Graphics are of interest to you.

In conclusion, after about 18 months as Program-Chair-Elect and Program Chair, I would like to thank Paul Murrell for organizing the data expo, Simon Urbanek as the 2006 Program-Chair-Elect (and thus responsible for the JSM 2007 in Salt Lake City - you may contact him at urbanek@research.att.com for questions regarding next year's JSM - any suggestions for 2007 are highly welcome!), all of our session organizers, chairs, speakers, discussants, and the ASA staff for their hard workin putting together this exciting program for 2006.

I wish you all the best - and have a great JSM 2006 in Seattle. See you soon!

Wendy Martinez (Section on Statistics in Defense and National Security - left) and Juergen Symanzik discussing the placement of a contributed paper during the February 2006 Program Chair Meeting in Alexandria, VA. (Photo courtesy of Juergen Symanzik)

Recent Meetings

FAST MANIFOLD LEARNING IN WILLIAMSBURG

Michael Trosset, 2006 Program Chair

This spring, the Section on Statistical Computing co-sponsored FAST MANIFOLD LEARNING, a 2-day workshop dedicated to the subject of scalable algorithms for nonlinear dimension reduction. Co-organized by two Section officers, Carey Priebe (Johns Hopkins University) and Michael Trosset (College of William & Mary), this event brought together 20 researchers from statistics, numerical analysis, and computer science, in order to share perspectives, learn about each other's work, and identify critical issues for future research. Other co-sponsors included the Interface Foundation of North America, the Office of Naval Research, AlgoTek Inc., and the College of William & Mary.

FAST MANIFOLD LEARNING was held in Williamsburg, VA, on April 14-15. Presentations included "DrLim: Dimensionality Reduction by Learning an Invariant Mapping" (Yann LeCun, Courant Institute of Mathematical Sciences), "Neighborhood Components Analysis" (Sam Roweis, University of Toronto), "Kernel Regularization and Dimension Reduction" (Fan Lu, University of Wisconsin), and "Non-Euclidean Multidimensional Scaling" (Jeffrey Solka, Naval Surface Warfare Center). Several presentations will be posted at http://www.math.wm.edu/~trosset/FML.

A sequel to FAST MANIFOLD LEARNING is being contemplated. The promotion of interdisciplinary exchanges on this subject is a natural mission for the Section on Statistical Computing. Statisticians have devised numerous techniques for the nonlinear dimension reduction of multivariate data sets. Recent research in machine learning has produced new techniques, collectively described as manifold learning. Although advances by the statistics and the machine learning communities have not yet been fully assimilated, intimate connections abound. Anyone interested in attending/organizing future events on this subject is encouraged to contact Carey Priebe <cep@jhu.edu> and/or Michael Trosset <trosset@math.wm.edu>.

COMPUTING SCIENCE AND STATISTICS: 38TH SYMPOSIUM ON THE INTERFACE

MASSIVE DATA SETS AND STREAMS

Yasmin H. Said
The Johns Hopkins University

California has long been symbolic of the American Dream, the culmination of the melting pot, containing a diverse mixture of cultures yet preserving the ideals of each. Pasadena, California fulfilled this reputation at the 38th Symposium on the Interface of Statistics, Computing Science, and Applications bringing together a mix of statisticians, mathematicians, computer scientists and applications experts both professional and student. Though the topic, Massive Data Sets and Streams, seemed narrow, by the first day of the conference it was clear that this theme was relevant in all fields as presentations ranged in subjects from heath-related issues to the acceleration of the expansion of the universe. Some presentations dealt with files that were hundreds of gigabytes in size while others dealt with just a single photon. The topic proved to be relevant in almost all aspects of modern society.

Upon arriving at the hotel, those registered for the conference were directed upstairs where they received a program of events, a nametag, a banquet ticket, several beverage tickets, and a wonderful tote bag featuring the emblem of the Interface. The Westin Hotel, which hosted the conference, was more than comfortable. Featuring a beautiful swimming pool with a hot tub and a lounging area that afforded a lookout over some of historic Pasadena, the hotel offered a many opportunities for relaxation. The gorgeous weather, mostly sunny with temperatures in the low eighties all week, made the pool even more inviting.

For those who wanted to explore California, the hotel is located just a few minutes walk from the center of Pasadena and the quaint historic downtown. A trip to In-N-Out Burger, essential for anyone planning to visit California was only a few blocks away as well. Within an hour's drive, Disneyland was a tempting thought for some. Others preferred the short drive into LA and exploring Hollywood. The Santa Monica Pier provided a breath-taking view of the beaches of the West Coast. A walk through Archieville, the center of Santa Monica with its immaculate streets featuring classy stores of all kinds, was the perfect way to spend an afternoon.

Sponsored by the Interface Foundation of North America, Inc. and financially by the National Security Agency, the ASA Sections on Statistical Computing and Statistical Graphics as well as the Jet Propulsion Laboratory, and SAS Inc., the conference started with a short courses on Wednesday morning. These two four-hour presentations made by experts in their fields were fascinating. The first presented by Dr. David Marchette dealt with the use of random graphs for pattern recognition. The second, presented by Professors Bin Yu and Mark Hansen, explained information theory and its relevance in statistics. These presentations provided an excellent start for the conference, inspiring interest in new techniques and methods. Throughout the day, drinks and snacks were served, which kept people awake and excited.

Represented at this conference were places all across the United States as well as some European countries and Australia. Because this Symposium was a joint conference with NASA, there were experts from many fields of science and mathematics. With such a diverse group in attendance, the mixer was a perfect event to break the ice. Less shy after a few drinks, many attendees were eager to talk about their projects and listen to others talk about their projects. Lining the walls of the room were colorful posters based. These poster presenters stood by their posters and enthusiastically explained them to passers-by. A genuine feeling of excitement was almost palpable in the air during the mixer. Attendees of the conference went to sleep with this excitement, eager for the next day of presentations.

On Thursday morning, Dr. Edward Wegman of George Mason University delivered the keynote speech. Usama Fayyad of Yahoo!, Inc. unfortunately could not deliver his planned keynote address on data mining. However, over two hundred guests attended Dr. Wegman's presentation about paleoclimate reconstruction and global warming. The "hockey stick" paleoclimate temperature reconstruction popularized in the climatology literature over the last eight years describes the temperature trends throughout the last millennium, with a sharp rise in present and future temperatures. This global temperature reconstruction has been challenged based on flaws in the mathematical methodology. While the documentary movie entitled An Inconvenient Truth featuring former Vice President Al Gore was to be released the day after Wegman's talk, Dr. Wegman pointed out an inconvenient truth about the movie itself, namely that the hockey stick prediction of rapidly rising temperature featured in the movie was based on a misuse of Principal Components Analysis. At the end of his talk Wegman with tongue in cheek gave his contact information for the Climate Police.

Amy Braverman and Yasmin Said

Interface Conference Chairs

Following the keynote address, the conference was in full throttle requiring five rooms for presentations. The program given to each guest upon arrival provided the abstracts for each paper and was organized by date, time, and subject. This program made it simple to find a presentation of interest. Since the presentations were designed to last anywhere from ten minutes to half an hour, it was easy to attend many of them, even without any previous knowledge of the subject material. In the morning, subjects ranged from information technology to genomics. As the day continued, a wide range of subject matter was covered during the conference including defense and security, astronomy and astrophysics, and issues in medicine. Also covered were subjects concerning telecom data streams, geophysics, and the analysis of data in higher dimensions.

After these presentations, guests were invited for a brief social gathering out on the sunny terrace that overlooked the hotel's beautiful fountains and once again brought the guests together to mingle not only as strictly professionals, but also as friends. During this time in the Fountain Room adjacent to the terrace, students again presented their posters as guests munched on fresh snacks. Furthering the bonds of friendship at the conference was the banquet following this mixer. Having no assigned seating at any of the presentations or conference events helped maintain the friendly and relaxed atmosphere throughout the conference. Also adding to this atmosphere was the lack of a dress code. Full suits, shirts and ties, collared shirts, or for some shorts and a t-shirt could all be found at the conference. As the event title "Banquet" suggested, this truly was a sumptuous feast. Fresh bread and salad was served to start, whetting the appetites of many a conference guest, hungry after the day of presentations. The main entrée was surf and turf, with ample portions of tender steak and fresh fish served with the choice of red or white wine. Sides of mashed potatoes and vegetables complemented the meat and fish. With food so tasty, guests struggled to save room for dessert. However, even those who thought they could not eat another bite had to try the dessert. The dessert for the evening was a chocolate pyramid. The only match for the beautiful presentation of the dessert was its delicious taste. The rich milk chocolate shell housed thick creamy chocolate mousse. After the food was served, the event continued as a jazz band filled the halls with delightful music.

Friday proceeded in much the same way as Thursday. Starting early in the morning, presentations of all kinds were given in five separate rooms. Subjects covered during Friday included challenges in data analysis, network traffic, space and solar physics, geoscience, forensic statistics, alcohol studies, and climate and weather. Following each presentation were brief question and answer sessions that were quite friendly and led to the clarification of many confusing points. These sessions served to achieve the underlying purpose of the Interface, to share knowledge and inspire interest in hundreds of subjects.

Saturday, the conclusion of the Interface, was only a half-day of presentations. Again starting early in the morning, topics, including computer models in national defense and homeland security, text mining, and data fusion, were covered.

Gathering at the Interface

Some highlights of the symposium included the sessions organized by David Scott and William Szewczyk on data streams. Of particular interest in these sessions was the description of new paradigms for streaming data involving computational environments where memory and processes on the processors are always fully utilized. Such an approach involves processing data to information without necessarily going through all of the intermediate steps that intuition would suggest, for example, understanding information in voice over IP without necessarily reconstructing the voice signal. In another fascination session on computational neuroscience organized by Dr. Emery Brown, Nichol Hatsopoulos discussed current developments in a cortically controlled brain machine interface. The focus on computational medical and neural science is a direction that has not been previously explored at the Interface Symposia. Another direction that has not been extensively explored is the exploitation of massive public health data sets. Dr. Paul Gruenwald, a first time participant at Interface presented a challenge to statisticians involving modeling of alcohol use systems from complex data systems. In a companion presentation, Professor William Wieczorek presented a discussion of the proliferation of massive datasets on alcohol and drug use epidemiology and how the analysis of these databases may offer a new opportunity for in-depth insights into these societal and individual problems.

A fortuitous connection to the last minute keynote on Statistics, Data Mining and Climate was the session Climate and Weather organized by Dr. Doug Nychka. This session featured a fascinating talk by Jeffrey Anderson on the assimilation of data and models to produce estimates of the state of the atmosphere. Dr. Nychka was a discussant in this session, which gave him an opportunity to comment on Dr. Wegman's keynote talk. The topic of forensic evidence has become a fascinating and very important new area of exploration. Other than DNA-related evidence, such traditional forensic evidence as polygraphs, bullet lead, fingerprints, and BAC measurements have or are coming under legal challenges. The Forensic Statistics session focused on the uniqueness of firearms-related evidence including the striations left in bullets by the muzzle of a gun. Benjamin Bachrach and James Filliben offered fascinating discussions. The ever-present concern with national security and homeland defense was reflected in a session organized by Professor David Banks. The session focused on traditional issues of validation and verification in a new modeling framework for homeland defense. (Professor Banks strong interest in homeland defense was reflected by his recent election as the Chair of the new ASA Section on Defense and National Security.)

A highlight of the Saturday morning was the session entitled Exploration of Massive Healthcare Databases, which was based on a paper competition from undergraduate students at The Johns Hopkins University. The students, who were taking their first course in statistics, tackled the 70-gigabyte HCUP dataset released by the Agency for Healthcare Research and Quality (AHRQ). HCUP is the Healthcare Cost and Utilization Project and consists of a database assembled by AHRQ from hospital reports all across the nation. Seeing undergraduate students so enthusiastic about statistics and medical research seemed to have the effect of revitalizing the senior statisticians as they remembered how excited they once were about their discipline.

The conference chairs Dr. Amy Braverman from the Jet Propulsion Laboratory and Dr. Yasmin Said from The Johns Hopkins University hoped that attendees of the Interface had a fruitful experience. The best feature of the conference was perhaps its relaxed nature. Guests were not obligated to attend anything that did not interest them. This sense of freedom allowed for the attendees to maximize the enjoyment of their trip to California. When it came time for checking out on Saturday, guests became sadly aware of how quickly three days could pass. These days were brimming with events, presentations and adventures. Conference goers drove or flew home on Saturday, many with work on Monday, with a rejuvenated interest in computing science and statistics.

Winners of a competition on the analysis of Healthcare Cost and Utilization Project (HCUP) data at Interface 2006. Ryan Archdeacon, Sophomore Biomedical Engineering major at The Johns Hopkins University, Eric Kim, Freshman Biomedical Engineering, Sarah Leismer Junior Biomedical Engineering, Mike Bisogno Sophomore Biomedical Engineering, Adam Sifuentes Junior Biomedical Engineering, Andrew Stirn Sophomore Electrical Engineering, Mathew Lalli Sophomore Biomedical Engineering, Devin O'Connor Junior Biomedical Engineering and Writing, and Derrick Kuan, Sophomore Biomedical Engineering at The Johns Hopkins University.

Statistical Computing Section Officers 2006

Stephan R. Sain, Chair
ssain@math.cudenver.edu
(303)556-8463
John F. Monahan, Chair-Elect
monahan@stat.ncsu.edu
(919)515-1917
Tim Hesterberg, Past-Chair
timh@insightful.com
(206)802-2319
Michael Trosset, Program Chair
trosset@math.wm.edu
(757) 221-2040
Ed Wegman, Program Chair -Elect
ewegman@gmu.edu
(703)993-1691
David J. Poole, Secretary/Treasurer
poole@research.att.com
(973)360-7337
Vincent Carey, COS Rep. 05-07
stvjc@channing.harvard.edu
(617) 525-2265
Robert Gentleman, COS Rep. 04-06
rgentlem@hsph.harvard.edu
(617) 632-5250
Juana Sanchez, COS Rep. 06-08 and Newsletter Editor
jsanchez@stat.ucla.edu
(310)825-1318
Thomas F. Devlin, Electronic Communication Liaison
devlin@mozart.montclair.edu
(973) 655-7244
Jose Pinheiro, Awards Officer
jose.pinheiro@pharma.novartis.com
(862) 778-8879
R. Todd Ogden, Publications Officer and Webmaster
ogden@cpmc.columbia.edu
212-543-6715
John J. Miller, Continuing Education Liaison
jmiller@gmu.edu
(703) 993-1690

Statistical Graphics Section Officers 2006

Paul R. Murrell, Chair and Electronic Communication Liaison
paul@stat.auckland.ac.nz
(649) 373-7599 x85392
Jeffrey L. Solka, Chair-Elect
jeffrey.solka@navy.mil
(540)653-1982
Mario Peruggia, Past Chair
peruggia@stat.ohio-state.edu
(614) 292-0963
Juergen Symanzik, Program Chair
symanzik@sunfs.math.usu.edu
(435) 797-0696
Simon Urbanek, Program Chair-Elect
urbanek@research.att.com
(973) 360 7056
John Castelloe, Secretary-Treasurer
John.Castelloe@sas.com
(919) 677-8000
Daniel B. Carr, COS Rep 05-07
dcarr@gmu.edu
(703) 993-1671
Edward J. Wegman, COS Rep 05-07
ewegman@galaxy.gmu.edu
(703) 993-1680
Naomi B. Robbins, COS Rep 04-06
naomi@nbr-graphs.com
(973) 694-2686
Dianne Cook, Newsletter Editor
dicook@iastate.edu
(515) 294 8865
Linda Williams Pickle, Publications Officer
picklel@mail.nih.gov
(301) 402-9344
Monica D. Clark, ASA Staff Liaison
monica@amstat.org
(703) 684-1221

Statistical COMPUTING & GRAPHICS

The Statistical Computing & Statistical Graphics Newsletter is a publication of the Statistical Computing and Statistical Graphics Sections of the ASA. All communications regarding the publication should be addressed to:

Dianne Cook, Editor Statistical Graphics Section. Department of Statistics Iowa State University, Ames, IA 50011-1210 (515) 294 8865 Fax: (515) 294 4040
dicook@iastate.edu
www.public.iastate.edu/-dicook

and

Juana Sanchez, Editor Statistical Computing Section. Department of Statistics University of California, 8125 MS Building, Los Angeles, CA90095 (310) 825-1218
jsanchez@stat.ucla.edu
www.stat.ucla.edu/-jsanchez

All communications regarding ASA membership and the Statistical Computing and Statistical Graphics Section, including change of address, should be sent to American Statistical Association, 1429 Duke Street Alexandria, VA 22314-3402 USA (703)684-1221, fax (703)684-2036 asainfo@amstat.org