The *WIND*

A newsletter from //Windward Technologies, Inc.

Q3 & Q4 97

INSIDE THIS ISSUE

Computer Breezes

- Intel develops a methodology to double the number of bits stored in the same space, as a single bit is currently stored. This discovery could allow a doubling of computer power every 9 months versus the current 18 months.
- Apple appoints co-founder Steve Jobs as acting CEO.
- Microsoft delays release of Windows 98. The release is now targeted for sometime between April and June of 1998.

Smoothed Survival Curves with ConFit

Donald E. Ramirez
der@Virginia.EDU
http://www.math.virginia.edu/~der/home.html

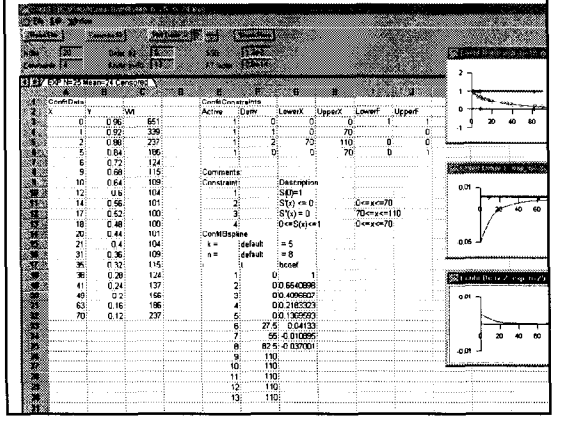
ConFit is a package designed to produce a smoothed approximating spline subject to a finite number of constraints on the function or its derivatives. The algorithm involves a least-squares problem which is solved for the coefficients of an appropriate B-spline. When constraints are imposed on the approximating spline, the problem is converted into a quadratic programming problem that **ConFit** solves with amazing speed.

I have used **ConFit** to calculate smoothed survival curves, and I will present an example below. I simulated 25 data values from an exponential distribution with a mean of 24. The largest 20% values were right-censored to model real-life data sets. The data used were {0, 1, 2, 5, 6, 9, 10, 12, 14, 17, 18, 20, 21, 31, 36, 38, 41, 49, 63, 70, 70+, 70+, 70+, 70+, 70+}. An estimate of the survival curve S(x) is given by the Kaplan-Meier method. This method also provides an estimate of the standard error for the estimates. The statistical software that I use is SPSS Release 7.0.

ConFit uses an Excel compatible spreadsheet for data entry and output (see Figure 1). I moved the data into the **ConFit** spreadsheet and chose the inverse of the square of the standard error for the weights in the least-squares regression. This follows the procedure for a logistic response function. See, for example, page 383 of Neter, J. *et al., Applied Linear Regression Models,* Richard Irwin, Inc., Homewood, Illinois (1983). The four constraints are:

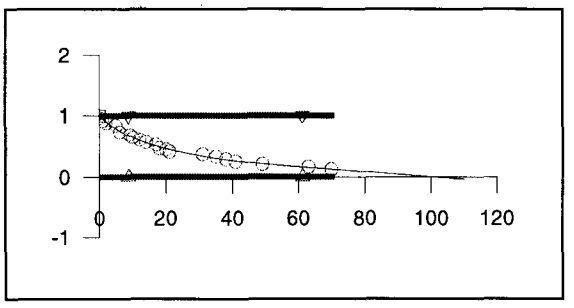
1. S(0) = 1
2. S′(x) <= 0 on [0,70]
3. S″(x) = 0 on [70, 120]
4. 0 <= S(x) <= 1 on [0, 70].

Figure 1: ConFit Environment

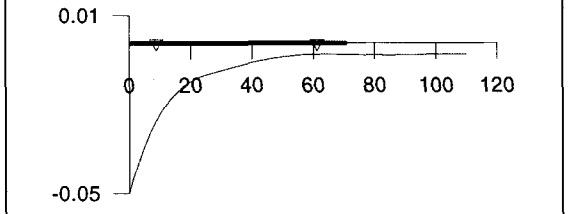


Constraint (2) forces S(x) to be monotone decreasing. Constraint (3) extends the natural spline linearly downward to the x-axis (see Figure 2).

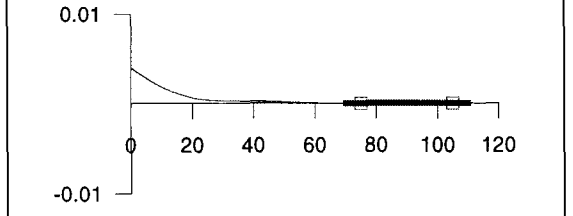
Figure 2: Linear extrapolation in ConFit Smoothed Kaplan-Meier Survival Curve



First Derivative of curve in Figure 2



Second Derivative of curve in Figure 2



I allowed **ConFit** to pick the spline by using the default values for the degree of the spline and the number of knots. The degree of the piecewise splines was five, and the number of knots used was 13. The Excel spreadsheet allowed me to easily transfer the knots and coefficients into Maple for further analysis. The calculated median was 17.1, with the true median being 16.6. When the number of data values are 100 or larger, I have gotten very good estimates for the quantiles. A plot of the hazard function h(x) = -S'(x)/S(x), using Maple, is shown in Figure 3.

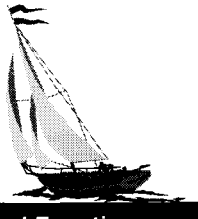
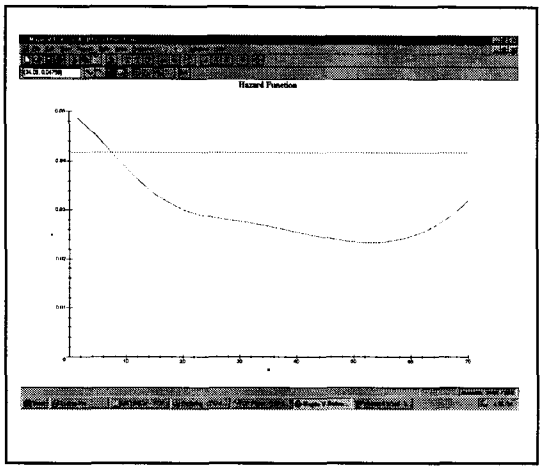
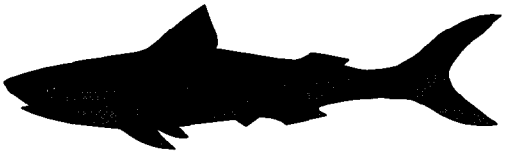


Figure 3: Hazard Function



To estimate the variance in the estimates for the quantiles, I propose resampling from the data repeating the procedure a number of times. This resampling technique is modeled after the smoothed bootstrapping method introduced in Gaylord, C. and Ramirez, D., Monotone regression splines for smoothed bootstrapping, *Computational Statistics Quarterly*, 6, 85-97 (1991).



What's New at WTI

WTI or its products can be found at the following Internet addresses:

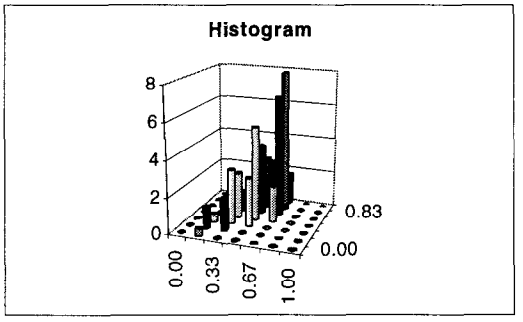
Visual Solutions (VisSim/OptimizePRO)
http://www.vissim.com/connect.htm

Mathworks (GRG2, BCLS, and RBFpack)
http://www.mathworks.com/connections/bcls_grg.shtml

Recovering Densities with RBFpack

RBFpack is a package (available in C and Fortran as well as a DLL) designed to fit **high** dimensional surfaces specified by **scattered** data. A **probability density** (density) is a function that is positive and integrates to one. Many times a density is approximated by a histogram that is based on a frequency count of observations over a given region. In two dimensions, such a histogram is usually based on frequency counts over rectangles. We have computed a histogram from experimental data and the plot follows.

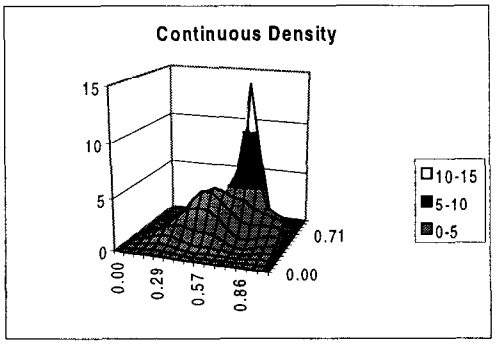
Initial Histogram Data



While this histogram might be adequate for many applications, it is true that continuous functions are preferred for many applications. This histogram has jump-discontinuities between the bins (rectangles) from which the frequencies are generated.

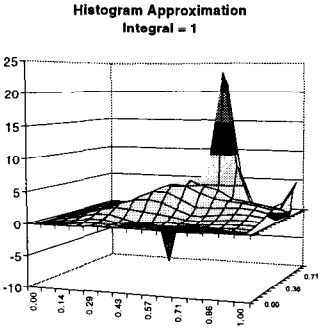
We can produce a continuous density which approximates the histogram by solving the following problem: Find the radial basis function f which minimizes the least squares distance from the histogram, which integrates to one, and which is nonnegative. Using the optimization routine GRG2 in combination with RBFpack allows an efficient solution to this problem. The resulting surface (with 10 centers) is:

Continuous density approximation
Integral = 1 and nonnegative



If we relax the conditions and just require the integral to be one, while not specifying that the fit must be positive. We get the following unsatisfactory surface approximation to the histogram (note the negative values).

Continuous approximation
Integral = 1
No constraints on values



Presentations and Publications

Phil Smith will give a talk on Splines in VBA (Visual Basic for Applications) at the University of Akron on October 10. The title is: Curve-fitting using VBA.

Tom Aird will attend the IFIP (International Federation for Information Processing) Working Group 2.5 business meeting and workshop on "Numerical and Programming Environment Aspects of DOE's ASCI (Advanced Scientific Computing Initiative) Modeling and Simulation Projects". The meeting is being held in Albuquerque, NM on October 16-19. The aim of IFIP WG 2.5 is to improve the quality of numerical computation by promoting the development and availability of sound numerical software. For more information about WG 2.5, consult the Web site:
http://www.nsc.liu.se/~boein/ifip

The *Wind* is a quarterly newsletter written and edited by Tom Aird and Phil Smith. Please let us know what your interests are and what sort of articles you would like to see. If you would like to contribute an article to the newsletter, please contact us at: WTI@aol.com or mail the sheet on page 4 back to us.

You can keep up with the latest information from ***WTI*** by visiting our Web site:
http://web.wt.net/~wti

// WindWard Technologies, Inc. http://web.wt.net/~wti

Products and Services

- **GRG2** Nonlinear optimization with bounds on variables and nonlinear constraints.
- **LSGRG2** Large-scale nonlinear optimization with bounds on variables and nonlinear constraints.
- **CURVI** Nonlinear optimization with bounds on variables.
- **BCLS** Linear least squares with bounds on variables and linear constraints.
- **RBFpack** Multidimensional scattered data fitting by radial basis functions.
- **VisSim/OptimizePRO** An add-on to VisSim a drag and drop simulation language.
- **MATLAB** interfaces for **GRG2**, **BCLS**, and **RBFpack**.

ON SITE CONSULTING We find that many of our clients have an explicit need to upgrade or add optimization to their products, but they do not have the time to invest in coming up to speed on the new technologies. At ***WTI*** we recognize this need and offer a site visit and consultation on your specific software needs. The cost for this service is $500 a day, plus expenses.

GUARANTEE We are so convinced in the quality of this service and our products in general that if you are not completely satisfied, we offer a 90 day money-back guarantee.

Let us hear from you!

By phone: 281-564-6523
By Fax: 281-564-6921
By E-mail: WTI@aol.com

Windward Technologies, Inc.
12039 Mulholland
Meadows, TX 77477

Windward Technologies, Inc.
12039 Mulholland
Meadows, TX 77477