Proceedings of the 1989 Winter Simulation Conference
E.A. MacNair, K.J. Musselman, P. Heidelberger (eds.)

A PRINCIPAL COMPONENTS MODEL OF SIMULATION STRUCTURE

David Alan Grier
Department of Statistics
George Washington University
Washington, DC 20052

1. OVERVIEW

It is often appropriate to analyze the output of a simulation study by looking at more than one measurement. In a traditional queuing model, it might be appropriate to look at three outputs:

1. the average time through the system for a customer,
2. the average number of balks per hour,
3. the average utilization of the servers.

These outputs measure different quantities which are usually related in their movements. Some pairs will be positively related, in that one variable will increase as the other variable increases. Likewise some will be negative related. The relationship can also be more complicated. For example, a recent simulation of a mining operation (Villanueva, 1989) measured, among other quantities, the amount of traffic in a mine tunnel leading to an ore shaft and the number of loads of ore removed through that shaft per hour. Initially, as the traffic increased, more loads of ore were removed. But as the tunnel became more and more crowded, the number of loads began to drop because of the traffic congestion.

Simulation models are often written to study nonlinear structure of the type described above, and it is useful to attempt a mathematical or graphical description of the relationships. This is particularly useful when simulation models are used to adjust the input parameters in order to optimize some particular output. The relationships between outputs can cause unintended side effects when one is adjusting the inputs of a model to optimize a single output. Attempts at optimizing one output while making an ad hoc attempt to control other outputs within a given range can be frustrating without a clear understanding of the relationships in the outputs.

The literature of multivariate statistics presents several techniques that can be used for exploring multivariate relationships in data. Examples include principal components and canonical correlation. As in much of traditional statistics, these techniques require that the data have gaussian distributions and that the relationships be linear. These assumptions limit their usefulness to simulation studies. In practice this means that the data have to be transformable in to something that is close to gaussian. Often simulation outputs are count data rather than continuous data. Their distribution frequently resemble asymmetric Poisson distributions rather than symmetric gaussians. Also, as demonstrated above, the relationships are often nonlinear.

To study the structure of a simulation model, we develop a principal components model of simulation output that is appropriate for count data. Let us start with some notation to describe the setting. The output of a simulation is data contained in a matrix

$$X = [x_{i,j}] \quad \begin{matrix} i = 1, \ldots N, \\ j = 1, \ldots, P. \end{matrix}$$

Each column of the matrix holds measurements on some given
item while each rows holds a collection of measurements taken
at a specific time. We will often refer to this model as a collection of P columns denoted $X_1, \ldots, X_p$. Using the above examples, the columns would represent variables such as the number of customers through the system or the number of balks. While in general, the model presented here does not require the rows (observations) to be independent the following treatment will assume that they are independent or nearly uncorrelated.

2. STANDARD TOOLS

In trying to put a structure on count data there are two tools that we shall combine: Poisson Regression and Principal Components. The traditional Principal Components analysis of a matrix attempts to find a linear structure in the data. This linear structure takes the form of P orthogonal vectors. The first of these vectors is the linear combination of the data columns which line which posseses the greatest amount of variability. The second of these combinations accounts for the second largest fraction of variability and so on. For example, suppose we have three columns in our matrix **X**, X_1, X_2, X_3 and the first principal component is
(0.24, 0.6, 0.6) and that this vector explains 60% of the variation. This direction implies that the three columns are positively related. The columns X_2 and X_3 move in the same amounts

whereas X_1 moves only 0.4 as much as the other columns.

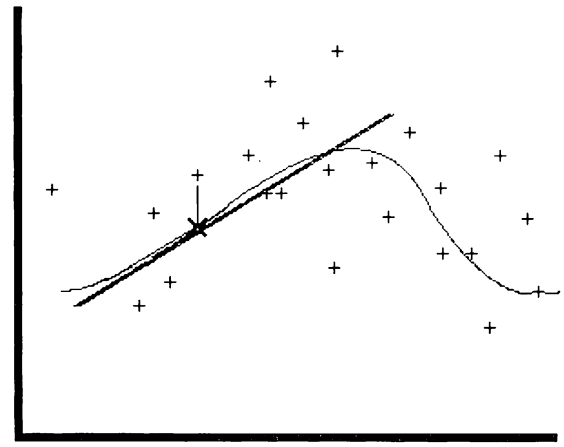
Principal components are computed by forming the eigenbasis of the sample covariance matrix. For gaussian data, the mean and the covariance matrix are enough to completely specify the distribution. However, that is the only multivariate distribution which may be specified by just the mean, first order structure, and the covariance matrix, second order structure. If the data are counts and each variable has a Poisson distribution, then there is no multivariate distribution that can be specified using just the means and the covariance matrix. Hence traditional principal components will have at best a dubious mathematical underpinning when applied to count data. We shall avoid this problem, in the sections below, by an algorithm that looks for nonlinear structure and operates directly on the data rather than on the covariance matrix.

Poisson Regression is a regression model for count data that is fit using maximum likelihood. Poisson Regression, attempts to fit linear structure to the expected value or intensity function of possion data. In the basic formulation of poisson regression there is a vector of n response variables Y and a matrix of explanatory variables **X**. Each of Y_i, i = 1, ..., n is an independent poisson random variable with intensity parameter $E(Y_i)$. The poisson regression model is $g(E(Y)) = \mathbf{X}B$ where g() is a link function which links the intensity parameters to the linear combinations of the explanatory variables. Note this is not an additive model, like the standard least squares regression model. There is no added poisson error. Instead the response variable has a poisson distribution with a intensity parameter that is a function of the covariates. The link function is usually chosen to be ln() and in that setting, the model becomes a multiplicative one.

Poisson regression is one member of the family of Generalized Linear Models (Nelder and Wedderburn, 1972). The parameters are fit using maximum likelihood and a modification of the least squares algorithm called iterated reweighted least squares. It is a very powerful tool, but it is not always the most appropriate tool for exploring the structure of a data set. It is common, as in our setting, to have P columns of data, X_1, ..., X_p, none of which is obviously the response vector.

3. SMOOTHING

Smoothing is a technique for analyzing structure in data that has recently received a great deal of attention (Buja, Hastie and Tibshirani, 1989). Given two vectors of data, X and Y, smoothing strives to fit a smooth function to data, Y = f(X). There are no constraints placed on f() other than it be smooth, that is f() is continuous and some number of derivations of f() are assumed to exist and be continuous. The exact estimate of the function depends on the algorithm to smooth the Y vector as a function of X. An easily understood smoothing technique is the running lines method. The vectors X and Y are sorted in parallel so that X is in ascending order. To find the value of the smooth function at some point x_i, a line is fit to the data in some bandwidth, say of length k, around x_i but at the same time excluding the point (x_i, y_i). The value of the function at x_i is the point on the line at x_i. This smoothing technique is illustrated in Figure 1.



Linear Line Smoothing
Figure 1

Besides running line smoothers, there are many techniques of smoothing including running means, smoothing splines, regression splines, running medians, and kernel density smoothers. Indeed, linear line fitting or standard linear regression can be viewed as a very restrictive form of smoothing. Smoothing, on the other hand, can be viewed as an extension of the standard regression methods. Indeed there is much current effort to extend all the tools of linear least squares regression to a nonparametric regression framework using smooths.

4. GAUSS-SEIDEL BACKFITTING ALGORITHMS

Smoothing is an efficient technique for estimating functions of a single variable but they rapidly become intractable for functions of more than one variable. The reason for this is the so-called "curse of dimensionality" (Friedman and Steutzle, 1981). This term refers to the problem that the number of points required to "fill" high dimensional space increases exponentially as the dimension increases. In the unit interval in one dimensional space, 10 equally spaced points are required if no two points are to be more than 0.1 units apart. In the unit hypercube in P space, 10^p points are required to meet the same condition. In practice, this means that in most reasonable data sets, there simply isn't enough data to get a good estimate of a function of P variables using a multivariate smoother.

One proposed solution (Friedman and Stuetzle, 1981) is the additive model. Instead of attempting to approximate the function $Y = f(X_1, X_2, \ldots, X_p)$, the estimation problem is restricted to functions of the form

$Y=f_1(X_1) + \ldots + f_p(X_p)$. In this setting, the functions can be estimated one at time, thus sidestepping the curse of dimensionality. The algorithm that does this fitting is called a "Backfitting Algorithm" or the "Gauss-Seidel Backfitting Algorithm" to indicate its relationship to the Gauss-Seidel Algorithm for solving linear equations. The algorithm estimates each of the $f_i()$ sequentially and then goes back and adjusts its estimates iteratively. This algorithm is given in Algorithm 1.

Algorithm 1: Gauss-Seidel Algorithm for fitting functions of the form
$Y = f_1(X_1) + \ldots + f_p(X_p)$

Step 0. Initialization

0.A Set $f_i() = 0.0$ for $i = 1, \ldots, P$

Step 1. Single Iteration

For $i = 1$ to P Do

1.A Set Y equal to $Y + f_i(X_i)$
1.B Smooth Y on X_i to produce a new estimate of $f_i(X_i)$
1.C Set Y equal to $Y - f_i(X_i)$

Step 2. Repeated Iteration

2.A Repeat Step 1 until convergence

The Gauss-Seidel algorithm is a simple but computationally intensive method for fitting models. While there is much practical experience to show that the estimates produced by the algorithm converge, there is only a little general theory that shows the estimates converge for any smoothers. It can be shown that the algorithm converges if the smoother is replaced by the conditional expectation operator $E(\ \ |X_i)$, which, in general, is unknown. The algorithm has been shown to converge for a class of linear shrinking smoothers which includes the spline smoothers (Buja, Hastie and Tibshirani, 1989).

The Gauss-Seidel Backfitting algorithm has been adapted to fit several different kinds of models including Generalized Additive Models (Hastie and Tibshirani, 1987), Principal Components (Donnell, 1987) and Principal Curves (Hastie and Stuetzle, 1989).

5. ALGORITHM FOR NONLINEAR PRINCIPAL COMPONENTS FOR COUNT DATA

Let X_i, $i = 1, \ldots, P$ be columns of data. Our intent is to find the functions $f_i(X_i)$, $i = 1, \ldots, P$, subject to the constraint that $f_i()$ is smooth for all i, so that the sum $f_1(X_1)+ \ldots + f_p(X_p)$ has the greatest possible variability. In practice, this will also maximize the pairwise correlations $Cor(f_i(X_i),f_j(X_j))$, $i <> j$.

The algorithm starts with the sum of all the variables and applies the backfitting algorithm. For each i, the variable X_i is subtracted from the sum and the result is smoothed on X_i to get the estimate of $f_i(X_i)$ which is most highly correlated with the remaining sum. This $f(X_i)$ is added to the sum and the process is repeated. After the smoothing process has been applied to each of the P variates, the entire process is repeated. The only difference is that now the current estimated of each function $f_i(X_i)$ is removed from the sum before the smoothing begins. In order to guarantee convergence, the $f_i()$ must be normalized. This can be done simply by scaling the f_i so that it has range 1.0. This restriction is also required in traditional Principal Components. In that setting, the norm of the coefficients of the linear combination is required to be 1.0. The algorithm is given formally in Algorithm 2.

Algorithm 2: Algorithm for Nonlinear Principal Components

0. Initialize

0.A Smooth X^*_i on the constant vector to get X_i

0.B Sum = $X^*_1+\ldots+X^*_p$
0.C $f_i = X_i$ $\quad i = 1, \ldots P$
0.D $SSQ_0 = ||$ Sum $||$
0.E J = 1
(number of current component)

1. Smooth
For $i = 1$ to p Do
1.A Sum = Sum - f_i
1.B Smooth Sum on X_i to produce a new f_i
1.C Normalize f_i to have range 1.0
1.D Sum = Sum + f_i

2. Iterate
2.A Repeat Step 2 until the f_i converge
2.B $SSQ_k = SSQ_{k-1} - ||$ Sum $||$

3. Find Additional Components
3.A J = J + 1
3.B Set $f_i = X_i$ for $i = 1, \ldots, P$
3.C Repeat steps 1 and 2

Algorithm 2 works well for exploring nonlinear structure that is similar to principal components in general data (Donnell, 1987). It may be applied to count data, although it can easily give misleading results when applied to Poisson data. The variance of a possion variable equals the mean, and hence comparing two variables with disparate magnitudes may be difficult. For example, suppose we have N observations of two Poisson random variables, X and Y. The expected values of X_i is $E(X_i) = f_i()$. For the purpose of this example let $f_i()$ be a function in the range [100,200]. If the expected value of X_j is small compared to $f_i()$, say $E(X_j) = 0.1$ f(i), it will be difficult to find any relationship between X_i and X_j, even when an

explicit relationship exists. In Poisson regression, this problem is handled by transforming the variables to the log scale. In that setting, the transformation to a log scale is motivated on the grounds that the sufficient statistics become linear combinations of the data. While sufficient statistics are not an issue in our setting and while it might be more mathematically defensible to use the square root transformation, which is variance stabilizing, we shall work with the log transformation. The main defense being the ease of interpretation of the components and the parallel with logistic regression.

6. SUMMARY

Principal Components is a useful tool for exploring multivariate data and the algorithm presented in this paper is an extension of that technique to include nonlinear modelling and count data.

REFERENCES

Brieman, L. and Friedman, J. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation" (with discussion), Journal of the American Statistical Association, 80, 580-619.

Buja, A., Hastie, T. and Tibshirani, R. (1989), "Linear Smoothers and Additive Models"(with discussion), Annals of Statistics, 17:2 (June 1989), 453-555.

Donnell, Deborah (1987), Nonlinear Principal Component Analysis, PhD dissertation, Department of Statistics, University of Washington, Seattle.

Friedman, J. and Steutzle, W., (1981), "Projection Pursuit Regression", Journal of the American Statistical Association,76, 817-823.

Friedman, Computationally Intensive Methods for Exploring and Modelling Multivariate Data, CBMS Monograph, Society for Industrial and Applied Mathematics, Philadelphia (to appear).

Hastie, T. and Tisbshirani, R., (1987), "Generalized Additive Models: Some Applications", Journal of the American Statistical Association, 82, 371-386.

Hastie, T. and Stuetzle, W.(1989), "Principal Curves", Journal of the American Statistical Association, 84:406 (June 1989), 502 - 516.

Morrison, D. F. (1976), Multivariate Statistical Methods, New York, McGraw-Hill.

Nelder, J. A., and Wedderbern, R.(1972), "Generalized Linear Models", Journal of the Royal Statistical Society, Ser. A, 135, 370-384.

McCullough, P., and Nelder, J. A.(1983), Generalized Linear Models, London, Chapman and Hall.

Villneuva, Manual, Study of the SIMSA Operations in San Ramon Peru, Master's Project, Department of Statistics, Computer and Information Science, The George Washington University, 1989.

AUTHOR'S BIOGRAPHY

DAVID ALAN GRIER is an assistant professor in the department of Statistics / Computer and Information Science at the George Washington University. He received a BA in mathematics from Middlebury College and a MS and PhD in statistics at the University of Washington, Seattle. His PhD work involved designing a system for simulation of statistical problems. Prior to his graduate study, he worked for Burroughs corporation doing software design and support for large scientific computers. His current research include computationally intensive methods of data analysis, numeric algorithms in artificial intelligence and simulation methods in statistics. He is a member of ASA, ACM and MAA.

David Alan Grier
Department of Statistics /
Computer and Information Science
Funger Hall
George Washington University
Washington, DC 20052