

PRACTICE OF COMPUTING DISTANCE-BASED REGRESSION - HOW MANY PC'S ARE RELEVANT?

Anna Bartkowiak, University of Wrocław
Przesmyckiego 20, 51-151 Wrocław, Poland.

Key Words: Prediction, Distance Matrix, Reduction of predictors

Abstract

Distance-based regression reduces finally to computation of ordinary LSE regression from principal coordinates. We compare the performance of two methods of assessing the importance of subsequent PC's in the final LSE regression: Cuadras' CP coefficient and the 'broken stick' rule. This is done by considering some real heliophysical data.

1. The distance-based regression

We are concerned with predicting the values of a variable Y from p explanatory variables (predictors) X_1, \dots, X_p , which can be of mixed type, i.e. continuous and/or categorical. We have observations of these variables for n individuals (items).

Cuadras and Arenas (1992) have proposed an interesting method, the distance-based regression (DBR), working as follows:

1. Firstly, a distance matrix \mathbf{D} of size $n \times n$ is evaluated from the explanatory variables (a variety of distances can be used here). The distance matrix \mathbf{D} is converted to the inner product matrix \mathbf{B} .
2. Next, some principal coordinates (PC's), denoted in the following as $\tilde{\Gamma}_1, \dots, \tilde{\Gamma}_m$, are constructed. The matrix \mathbf{B} is decomposed into rank one matrices built from principal coordinates:

$$\mathbf{B} = \sum_{i=1}^m \lambda_i \Gamma_i \Gamma_i^T = \sum_{i=1}^m \tilde{\Gamma}_i \tilde{\Gamma}_i^T, \quad (1)$$

with $\Gamma_i^T \Gamma_j = \delta_{ij}$, $\tilde{\Gamma}_i = \lambda_i \Gamma_i$.

3. Finally, an ordinary LSE regression is evaluated from the established PC's. Usually, a reduced number $k < m$ of all the PC's is taken into consideration:

$$Y = \beta_0 + \beta_1 \tilde{\Gamma}_{i_1} + \dots + \beta_k \tilde{\Gamma}_{i_k} + \epsilon_{(k)} \quad (2)$$

2. The number of relevant PC's

How many PC's can be obtained from the inner product matrix \mathbf{B} ?

Let us suppose that the recorded data do not reveal any linear dependence among values of the considered predictors, i.e. among the columns of the data matrix. Moreover, suppose that $n > p$, and all the data vectors for the recorded individuals are different. Data satisfying these conditions will be referred to as one being in standard conditions.

Now suppose that our recorded data are in standard conditions, and we calculate from them the matrix \mathbf{B} and the decomposition given by eq. (1).

Using Euclidean distances we obtain under standard conditions p PC's. Then, the regression defined by (2) yields exactly the same multiple correlation coefficient and residuals as the ordinary classical LSE method. Since the last method is much easier and faster to calculate, there is no need and no advantage to use the DBR in that case.

Using L1-norm or Gower distances we obtain under standard conditions $m = n - 1$ PC's. Moreover, we obtain a complete explanation of Y by the derived m PC's, what is due to overfitting of the model. Therefore a smaller number $k < m$ of all the PC's should be taken into consideration.

Usually we put into the regression (2) those PC's which are mostly correlated with Y . It is known that they are not exactly the *first* k PC's, i.e. those connected with the k largest eigenvalues of \mathbf{B} . So, to be safe, we should compute *all* the eigenvectors of \mathbf{B} and next choose those with the highest correlation with Y . On the other hand, computing all the m PC's is cumbersome and much inconvenient for large values of n .

To assess the relevance of subsequent PC's in (2) we could perform a statistical test of significance, e.g. a F test. However, this is questionable, because statistical tests used for this purpose need assumption of normality,

what is again doubtful in the case of DBR, especially when dealing with mixed type of predictors.

In the following we will be concerned with two aspects of choice of the relevant PC's:

1. Looking for a cutting rule allowing to state at some stage of decomposition of \mathbf{B} (formula 1), that all the relevant PC's are already obtained;
2. Finding a distribution-free substitute of statistical test allowing to judge the relevance of the extracted PC's.

Cuadras et al. (1993) have proposed an empirical procedure based on the defined by them *CP coefficient of predictability*.

We propose another empirical procedure using the *broken stick* rule.

In the following we will explain in more detail the two mentioned empirical procedures of selection of a smaller number of PC's. We will apply them to the data described by Bartkowiak & Jakimiec (1994), or Jakimiec & Bartkowiak (1994), and compare the number of PC's indicated as relevant by these two methods.

3. CP, the coefficient of predictability

The coefficient of predictability, introduced by Cuadras et al. (1993), is defined as follows:

$$CP(i) = \frac{\mathbf{y}^T \mathbf{B}_{(i)} \mathbf{y}}{\mathbf{y}^T \mathbf{B} \mathbf{y}}, \quad i = 1, \dots, m, \quad (3)$$

with $\mathbf{B}_{(i)} = \mathbf{B} - \sum_{j=i+1}^m \lambda_j \Gamma_j \Gamma_j^T$ and \mathbf{y} denoting the $n \times 1$ vector of observed values of the variable Y .

The denominator of (3) can be decomposed as:

$$\mathbf{y}^T \mathbf{B} \mathbf{y} = \sum_{j=1}^m r_j^2 \lambda_j,$$

with r_j being the Pearsonian correlation coefficient between \mathbf{y} and $\tilde{\Gamma}_j$. Obviously $0 < CP(i) \leq 1$, $CP(m) = 1$, $CP(i) \leq CP(i+1)$.

Cuadras et al. have proposed to watch the diminishing of $CP(i)$ when subsequent PC's are extracted from \mathbf{B} . The PC's that diminish

Table 1: Values of $1 - CP(i)$ indicating the unexplained part of $Y^T B Y$ when working with L1-norm and Gower distances

L1-norm dist.			Gower dist.		
i	Y=Mv	Y=Fs	i	Y=Mv	Y=Fs
1	.0770	.0269	1	.0937	.0301
2	.0355	.0260	2	.0355	.0253
3	.0355	.0260	3	.0355	.0252
4	.0342	.0248	4	.0341	.0240
5	.0341	.0248	5	.0341	.0237
6	.0193	.0103	6	.0200	.0104
7	.0184	.0098	7	.0191	.0097
8	.0184	.0098	8	.0190	.0097
9	.0131	.0073	9	.0141	.0078
10	.0131	.0073	10	.0138	.0075
11	.0127	.0072	11	.0120	.0068
12	.0109	.0055	12	.0111	.0057
13	.0103	.0055	13	.0107	.0057
14	.0098	.0054	14	.0100	.0056
15	.0098	.0050	15	.0100	.0051

the values of CP relatively 'much', are supposed to be 'relevant' in prediction of Y .

The quantity $1 - CP(i)$ shows the amount of the denominator in (3) that is not explained yet. Cuadras et al. have proposed to watch the graph of $1 - CP(i)$ as put against i . If the abscissa is near 0, then all relevant information on predicting Y is already accounted for and we are justified to stop the process of extracting the further PC's.

We have applied this method to the heliophysical data described in the paper of Bartkowiak and Jakimiec (1994), hereafter called *B&J*. They tried to predict two variables, $Y1 = Mv$ and $Y2 = Fs$ denoting: Mv - the maximum value of solar flare X-ray flux (fs), and Fs - the total sum of the fs amount. Similarly as *B&J*, we consider here also Euclidean, L1-norm and Gower distances. The values of $1 - CP(i)$ for Euclidean distances are shown in the last column of Table 2, the respective values for L1-norm and Gower distances are shown in Table 1. The values of $1 - CP(i)$ that differ from $1 - CP(i-1)$ by more than 0.01 are underlined (it is assumed that $1 - CP(0) \equiv 1.00$).

One can see that in all cases the first PC is the most relevant: it reduces the initial value of CP by more than 90%. The impact of the

Table 2: Euclidean distances. Eight highest r_i^2 , limits (LO, UP) from *broken stick* rule, and values $1 - CP(i)$; h denotes original id number of the PC

i	r_i^2	h	LO	UP	1-CP(i)
Predicted: Y=Mv					
1	*.293	1	.095	.161	.0288
2	*.044	5	.064	.099	.0133
3	.015	2	.044	.072	.0125
4	.014	4	.030	.053	.0079
5	.010	7	.020	.041	.0005
6	.002	3	.012	.030	.0005
7	.000	6	.006	.020	.0000
8	.000	8	.001	.011	.0000
Predicted: Y=Fs					
1	*.418	1	.121	.208	.0122
2	.032	5	.083	.126	.0114
3	.026	4	.057	.091	.0104
4	.003	3	.039	.069	.0041
5	.003	7	.025	.052	.0001
6	.001	8	.014	.037	.0001
7	.001	2	.006	.025	.0000
8	.000	6	.001	.014	.0000

remaining PC's is really a small one. It can be seen that the results obtained for the L1 norm and Gower distances are very similar (this has been already pointed out by B&J).

The first PC constructed from Euclidean distances allows for reconstruction of $\mathbf{y}^T \mathbf{B} \mathbf{y}$ in more than 97% both for $Y = Mv$ and $Y = Fs$; analogous first PC from L1 norm or Gower distances makes this in more than 90% for Mv and about 97% for Fs.

The second PC seems to have an impact only when considering $Y = Mv$ and using L1 norm or Gower distances.

The 6th PC seems to be relevant for both variables, however only when working with L1 norm or Gower distances.

4. The broken stick rule.

Suppose, we have a stick of unit length, which is broken, at random, into p segments. Then it can be shown that the expected length l_i of the i th longest segment is (ctf. Jolliffe 1986, p. 95):

$$l_i = \frac{1}{p} \sum_{j=i}^p \frac{1}{j}. \quad (4)$$

Table 3: L1-norm distances. Fifteen highest r_i^2 's and limits (LO, UP) from *broken stick* rule; h denotes original id number of the PC

i	r_i^2 Mv	h	r_i^2 Fs	h	LO	UP
1	*.279	1	*.416	1	.033	.051
2	*.043	36	*.054	6	.029	.040
3	*.042	43	.024	80	.027	.034
4	*.039	2	.022	43	.025	.031
5	*.039	6	.022	62	.023	.029
6	*.029	9	.020	12	.022	.027
7	*.027	51	.019	9	.021	.026
8	*.026	46	.018	52	.020	.024
9	.021	21	.018	36	.019	.023
10	.019	24	.018	24	.019	.022
11	.018	62	.017	84	.018	.021
12	.017	57	.017	21	.017	.020
13	.016	83	.016	29	.017	.019
14	.015	12	.015	111	.016	.019
15	.014	70	.014	46	.016	.018

The standard deviation of the i th longest segment can be obtained by simulation. We have used 2000 repetitions.

We have applied this rule to our data when considering the squared correlations r_i^2 , $i = 1, \dots, m$ between the vector \mathbf{y} and the constructed principal coordinates $\tilde{\Gamma} = (\tilde{\Gamma}_1, \dots, \tilde{\Gamma}_m)$. Using Euclidean distances we got $m = 8$ PC's, using L1-norm or Gower distances we got $m = 129$ PC's.

Since the constructed PC's are mutually uncorrelated, the total squared multiple correlation coefficient R^2 between \mathbf{y} and the set $\tilde{\Gamma}$ can be decomposed as

$$R^2(\mathbf{y}, \tilde{\Gamma}) = \sum_{i=1}^m r_i^2. \quad (5)$$

Taking this into account we apply the broken stick rule to the ordered r_i^2 's. The results, i.e. the ordered r_k^2 's and their respective one-sigma lower and upper limits obtained from the assumption of a random subdivision - are shown in Table 2 (for the Euclidean distances), Table 3 (for the L1-norm distances) and Table 4 (for the Gower distances).

When working with Euclidean distances we got $R^2(\mathbf{y}, \tilde{\Gamma})$ equal to 0.3800 and 0.4846 for $Y = Mv$ and $Y = Fs$, respectively. Then we have assumed, that we have to deal with

Table 4: Gower distances. Fifteen highest r_i^2 's and limits (LO, UP) from *broken stick* rule; h denotes original *id* number of the PC

i	r^2 Mv	h	r^2 Fs	h	LO	UP
1	*.266	1	*.413	1	.041	.043
2	*.055	2	*.052	6	.034	.035
3	*.038	6	.024	22	.030	.031
4	*.037	46	.022	46	.028	.028
5	*.035	36	.021	84	.026	.026
6	*.027	9	.020	61	.024	.025
7	*.026	22	.019	57	.023	.024
8	.022	43	.018	36	.022	.022
9	.019	24	.018	81	.021	.021
10	.017	57	.015	9	.020	.020
11	.016	73	.014	124	.019	.020
12	.015	51	.014	111	.019	.019
13	.015	44	.013	12	.018	.018
14	.014	84	.013	80	.017	.018
15	.013	11	.013	29	.017	.017

sticks of 0.3800 and 0.4846 length broken into 8 parts.

When working with L1-norm or Gower distances we have $R^2(\mathbf{y}, \tilde{\Gamma}) = 1$, a strict linear dependence for both $Y = Mv$ and $Y = Fs$, what means that putting into equation (2) all the m PC's we are able to predict accurately the values of Y , i.e. with 0 error. This fantastic result due to overfitting is a spurious one and is true only when making self validation, i.e. when predicting the values of Y in the same data set from which the PC's were evaluated. When making predictions in a foreign data set the situation might be quite different. What we really want - is to find the essential PC's that describe the model of the analysed data. We suppose that those that exceed in magnitude the segments of the broken at random stick might have this property.

Let us look now at Tables 2, 3 and 4. We have assumed that values which exceed the one-sigma bound (LOW,UPP) can be judged as *signicative*, i.e. coming from a non-random subdivision of the total. Values satisfying that condition are marked by an '*' sign.

One can see, that using Euclidean distances, only the first PC, yielding a correlation coefficient $r_1^2 = 0.293$ for $Y = Mv$ and $r_1^2 = 0.418$ for $Y = Fs$, appears to be significant.

Using L1-norm and Gower distances we

obtain 7-8 significant r_i^2 's for $Y = Mv$ and two significant r_i^2 's for $Y = Fs$.

Thus, it appears that for $Y = Fs$ both methods (i.e. the *CP* method and the *broken stick* method) indicate the same signicative PC's.

5. Discussion of the results

The *CP* rule is much more economical than the 'broken stick' method, because subsequent PC's are extracted stepwise, i.e. when required. For our data certainly not more than 15 PC's (from total 129) are needed. However, the stopping rule is based on some subjective judgment. It happened in our data (with p variables) that the first 8 PC's reduce the initial value of the criterion by about 98%.

The broken stick rule seems to indicate more PC's that are truly relevant. This method is more difficult to carry out, because the computing of all PC's is needed. The gain is, that it specifies an objective rule of choosing the relevant PC's.

It remains to the purview of the user to make the proper choice.

Acknowledgments

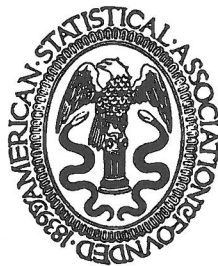
I express my cordial thanks to prof. Maria Jakimiec for reading this manuscript and giving her comments.

References

- Bartkowiak A. & Jakimiec M. (1994): Distance-based regression in prediction of solar flare activity. *Qüestiió* 18, 271-300.
- Cuadras C.M., Arenas C. & Fortiana J.G. (1993). Further aspects of a distance based model for prediction including non linear regression. *LINSTAT'93*, Poznań (Poland), 1-4 June 1993.
- Cuadras, C.M. & Arenas, C. (1990). A distance based regression model for prediction with mixed data. *Commun. Statist. Theory Meth.*, 19, 2261-2279.
- Jakimiec M., & Bartkowiak A. (1994): Short-term Solar Flare Predictions by Distance-based Regression I. Bearalert Regions in 1988 and 1989 - Continuous Predictors. *Acta Astronomica* 44, 115-140.
- Jolliffe I.T. (1986). *Principal Component Analysis*. Springer New York, Berlin.

American Statistical Association

1994
Proceedings
of the
Statistical Computing Section



Papers presented at the
Annual Meeting of the American Statistical Association,
Toronto, Canada, August 13-18, 1994,
under the sponsorship of the
Statistical Computing Section

American Statistical Association ■ 1429 Duke Street, Alexandria, VA 22314

TABLE OF CONTENTS

Invited Papers by Topic

I. Estimation for Stochastic Processes Through Simulation

Organizer: *Katherine B. Ensor*, Rice University

Chair: *Monnie McGee*, Rice University

Inference for Autocorrelations by Resampling.

Joseph P. Romano, Stanford University; *Lori A. Thombs*, University of South Carolina 1

II. Neural Network Models

Chair/Organizer: *Trevor Hastie*, AT&T Bell Laboratories

Using Neural Networks to Learn Intractable Generative Models.

Geoffrey E. Hinton, *Peter Dayan*, *Radford M. Neal*, and *Richard S. Zemel*, University of Toronto 11

III. Self-Validating Numerical Methods in Statistical Computing

Chair/Organizer: *William J. Kennedy*, Iowa State University

Applications of Numerical Interval Analysis to Obtain Self-Validating Results
in a Massively Parallel Computing Environment.

Ouhong Wang, Iowa State University 21

An Object Orient Computing Environment for Self-Validating Computing.

Morgan C. Wang, University of Central Florida 29

Error-Free Least Squares Based on Multiple Homomorphic Images.

Sallie Keller-McNulty, Kansas State University 33

IV. Computationally Intensive Statistical Methods

Chair/Organizer: *James E. Gentle*, George Mason University

Cross-Validation, the Bootstrap, and Related Methods for Tuning Parameter Selection.

Naomi Altman, Cornell University; *Christian Léger*, University of Montreal. 41

Robust Sample Survey Inference via Bootstrapping and Bias Correction: The Case of the Ratio Estimator.

R.L. Chambers, Australian National University; *A.H. Dorfman*, U.S. Bureau of Labor Statistics 51

V. The Frontiers of Maximum Entropy

Chair/Organizer: *Harry F. Martz*, Los Alamos National Laboratory

Applications of Quantum Entropy to Statistics.

R. N. Silver and *H. F. Martz*, Los Alamos National Laboratory 61

Beyond Maximum Entropy.

John Skilling and *Sibusiso Sibisi*, University of Cambridge 71

VI. Hidden Markov Models

Chair/Organizer: *Karen Kafadar*, University of Colorado

- Hidden Markov Models and Their Estimation. *Arthur Nádas*, New York University;
Robert L. Mercer, Renaissance Technologies, Inc. 80
- Topics on Hidden Markov Models and Their Applications in Speech Recognition.
Don X. Sun, State University of New York; *Li Deng*, University of Waterloo;
C.F.J. Wu, University of Michigan 90

Contributed Papers by Topic

I. Smoothing and Maximum Likelihood Methods

Chair: *Kevin Chartier*, Kansas State University

- Adaptive Signal Regression.
Stephanie R. Land and *Jerome H. Friedman*, Stanford University 100
- Automatic Smoothing Parameter Selection for Robust Nonparametric Regression.
Ferdinand T. Wang, American University 106
- On the Optimality of Prediction Based Selection Criteria and the Convergence Rates of Estimators.
Naomi Altman, Cornell University; *Christian Léger*, University of Montreal 112
- Influence on Smoothness in Penalized Likelihood Regression for Binary Data.
Robert W. Jernigan, The American University; *Julie A. O'Connell*, Westat, Inc. 118
- The Profile Z Plot and Confidence Intervals in Maximum Likelihood Analysis.
Jian-Shen Chen, Chaoyang Institute of Technology;
Robert I. Jennrich, University of California, Los Angeles 124
- Permuting CM Steps Within the ECM Algorithm.
David A. van Dyk and *Xiao-Li Meng*, University of Chicago 130

II. Systems: Belief, Database, ASSIST, Operating, and Algebraic

Organizer: *Vicki Lancaster*, Kansas State University

Chair: *Linda A. Clark*, AT&T Bell Laboratories

- Dealing with Uncertain Knowledge in Knowledge Base Systems.
Morteza Marzjarani, Saginaw Valley State University 136
- Advantages of Developing, Testing and Accessing Complex Statistical Tools in an ASSIST Environment.
Marcus John Sanchez, National Center for Health Statistics 141
- Computer Algebra in Statistics: Some Examples.
G.U.H. Seeber, Leopold-Franzens University 147

III. Regression: Model Selection and Assumption Violation

Chair: *Mark S. McNulty*, Kansas State University

- Partially-Ordered Linear Models.
John H. Walker, Cornell University 152
- Model Selection in Tree-Structured Regression.
Thomas W. Miller, University of Oregon 158

Computation of Determinantal Subset Influence in Regression. <i>Bruce E. Barrett and J. Brian Gray, University of Alabama</i>	164
A Likelihood Ratio Type Test for Heteroscedasticity. <i>Yogendra P. Chaubey, Concordia University; Enayet Talukder, Otsuka America Pharmaceutical, Inc.</i>	170
Durbin-Watson Revisited. <i>Michael C. Axelrod and Jeffrey G. Glosup, Lawrence Livermore National Laboratory</i>	176
IV. Distribution Estimation: Numerical Integration, Rank Procedures, Bootstrapping, and Random Number Generation Chair: Morteza Marzjarani, Saginaw Valley State University	
Random Integration Rules for Statistical Computation. <i>Alan Genz and John Monahan, North Carolina State University</i>	182
On the Normal Approximation for the Sample p-Quantile. <i>Kallappa M. Koti, Pennsylvania State University</i>	188
Testing Whether a Count Variable Belongs to the Katz Family of Distributions. <i>Kevin J. Anstrom and Dana Quade, University of North Carolina</i>	194
Rank Tests With Weighted Data. <i>Robert J. Blodgett, U.S. Food and Drug Administration</i>	200
A Comparison of Bootstrap Approaches for Evaluating Dollar Unit Samples in Auditing. <i>Howard R. Clayton, University of North Texas</i>	205
V. Comparison of Experimental Design and Analysis Methods Chair: Marie M. Loughin, Applied Research Consultants	
When is Satterthwaite's Approximation Adequate? <i>Michael D. Conerly, University of Alabama; J. Michael Hardin, University of Alabama at Birmingham; Lance A. Waller, University of Minnesota</i>	209
Applicability of Nine Numerical Techniques for Detecting Active Factors in Unreplicated Experimental Designs <i>Claudio Benski, Schneider Electric</i>	214
Statistical Analysis Appropriate for K Samples of Two Variables. <i>Brahmanand Nagarsenker and Panna Nagarsenker, Air Force Institute of Technology</i>	218
Issues Related to the Bioequivalence of Drug Formulations. <i>Mohamed Habibullah, University of Wisconsin; James Rogers, Research & Data Services, Inc.; Theresa Voss, Clintec Nutrition Company</i>	222
VI. Messy Data: Linear, Nonlinear, and Multivariate Models Organizer: Vicki Lancaster, Kansas State University Chair: Dallas E. Johnson, Kansas State University	
Multiplicative Effects in Three-Way Analysis of Variance. <i>K. See, Miami University; Eric P. Smith, Virginia Polytechnic Institute and State University</i>	225
A Comparison of Two SAS [®] Procedures in Decomposing the Structure of Two-Way Tables. <i>Bahman Shafii, University of Idaho</i>	231

Practice of Computing Distance-Based Regression-How Many PC's are Relevant? <i>Anna Bartkowiak</i> , University of Wroclaw	235
Heteroscedastic Nonlinear Regression When There Are Possible Outliers <i>László Tóthfalusi and László Endrényi</i> , University of Toronto	239
Non-linear Time Series Models for Ordinal Data. <i>Raymond G. Hoffmann</i> , The Medical College of Wisconsin	245
VII. Classification/Clustering	
Organizer: <i>Kenneth T. Higbee</i> , Battelle, Richland	
Chair: <i>William F. Szewczyk</i> , National Security Agency	
A New Metric For Data Clustering. <i>David B. Millar</i> , U.S. Department of Defense	251
Estimation of Within-Group Covariance. <i>Jeffrey L. Townsend</i> , U.S. Department of Defense	257
Variably Regularized Discriminant Analysis. <i>Kenneth T. Higbee</i> , Battelle, Northwest	263
Contributed Papers—Poster Sessions	
The Software Taxi: An Easy-to-Use Tool for Statistical Presentations. <i>John C. Nash</i> , University of Ottawa	269
SRS Runs and Spacings Tests to Assess Randomness. <i>W. D. Kaigh and E. F. Schuster</i> , The University of Texas at El Paso	272
A Comparison of Power Approximations for Satterthwaite's Test. <i>Rachael L. DiSantostefano</i> , Family Health International; <i>Keith E. Muller</i> , University of North Carolina	278
Parameters and Quantiles Estimation for Continuously Distributed Random Variables. <i>Enrique Castillo</i> , University of Cantabria; <i>Ali S. Hadi</i> , Cornell University	284
Solving the Ranking and Selection Indifference-Zone Formulation for Normal Distributions Using Computer Software. <i>Catherine A. Poston, David R. Barr and Paul F. Auclair</i> , Air Force Institute of Technology	290
Analysis of Square Tables With Ordered Categories. <i>Wai-Yin Poon and Hin-Yan Hung</i> , The Chinese University of Hong Kong	292
Index	299