Scan Statistics for Normal Data with Applications

Joseph Glaz

University of Connecticut

October 2013

Outline of the Presentation

- Scan statistics for normal data.
  - Introduction
  - Probability Inequalities: one dimensional data.
  - Product-type approximations for one dimensional data.
  - Inequalities and approximations: two dimensional data.
  - Variable window scan statistics: minimum p-value approach.
  - Applications to time series data.
- Summary and future work.

Introduction: Early References on Clustering of Events

- Berg, W. (1945). Aggregates in one-and-two-dimensional random distributions. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **36**, 337-346.
- Dinneen, G. P. and Reed, I. S. (1956). An analysis of signal detection and location by digital methods. *IRE Trans. Information Theory*, **IT**-**2**, 29-39.
- Domb, C. (1950). Some probability distributions connected with recording apparatus II. *Proceedings Cambridge Phil. Soc.*, **46**, 429-435.
- Mack, C. (1948). An exact formula for $Q_k(n)$, the probable number of k- aggregates in a random distribution of n points. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **39**, 778-790.
- Silberstein, L. (1945). The probable number of aggregates in random distributions of points. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **36**, 319-336.

Introduction: Early Theoretical Advances on Scan Statistics

- Naus, J. I. (1963). *Clustering of Random Points on the Line and Plane*. P.D. Thesis, Harvard University, Cambridge, MA.
- Ikeda, S. (1965). On Bouman-Velden-Yamamoto's asymptotic evaluation formula fro the probability of visual response in a certain experimental research in quantum biophysics of vision. *Ann. Inst. Stat. Math*. **17**, 295-310.
- Naus, J. I. (1965a). The distribution of the size of the maximum cluster of points on a line. *J. Amer. Stat. Assoc*. **60**, 532-538.
- Naus, J. I. (1965b). Power comparison of two tests of non-random clustering. *Technometrics* **8**, 493-517.
- Barton, D. E. and Mallows, C. L. (1965). Some aspects of the random sequence. *Annals of Mathematical Statistics* **36**, 236-260.
- Leslie, R. T. (1969). Recurrent times of clusters of Poisson points. *J. Applied Probability* **6**, 372-388.

Introduction: Early Theoretical Advances on Scan Statistics

- Cressie, N. (1977). On some properties of the scan statistic on the circle and the line. *J. Applied Probability* **14**, 272-283.
- Cressie, N. (1980). The asymptotic distribution of the scan statistic under uniformity. *Annals of Probability* **8**, 838-840.
- Glaz, J. (1979). Expected waiting time for the visual response. *Biological Cybernetics* **35**, 39-41.
- Glaz, J. and Naus J. (1979). Multiple coverage of the line. *Annals of Probability* **7**, 900-906.
- Greenberg, I. (1970). The first occurrence of N successes in N trials. *Technometrics* **12**, 627-634.
- Huntington, R. J. and Naus, J. I. (1975). A simpler expression for the Kth nearest neighbor coincidence probabilities. *Annals of Probability* **3**, 894-896.

Introduction: Early Theoretical Advances

- Huntington, R. J. (1978). Distribution of the minimum number of points in a scanning interval on the line. *Stochastic Processes and their Applications* **25**, 73-77.
- Hwang, F. K. (1974). A discrete clustering problem. *Manuscript*, Bell Labs, Murray Hill.
- Hwang, F. K. (1979). A generalization of the Karlin-McGregor theorem on coincidence probabilities and an application to clustering. *Annals of Probability* **5**, 814-817.
- Karlin, S. and McGregor, G. (1959). Coincidence probabilities. *Pacific Journal of Mathematics* **9**, 1141-1164.
- Naus, J. I. (1974). Probabilities for a generalized birthday problem. *J. Amer. Stat. Assoc.* **69**, 810-815.
- Naus, J. I. (1979). An indexed bibliography of clusters, clumps, and coincidences. *International Statistical Review* **47**, 47-

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
  - Agricultural Sciences

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
  - Agricultural Sciences
  - Astronomy

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
    - Agricultural Sciences
    - Astronomy
    - Bioinformatics

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
  - Agricultural Sciences
  - Astronomy
  - Bioinformatics
  - Biosurveillance

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
  - Agricultural Sciences
  - Astronomy
  - Bioinformatics
  - Biosurveillance
  - Ecology and Environmental Sciences

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
  - Agricultural Sciences
  - Astronomy
  - Bioinformatics
  - Biosurveillance
  - Ecology and Environmental Sciences
  - Epidemiology

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
  - Agricultural Sciences
  - Astronomy
  - Bioinformatics
  - Biosurveillance
  - Ecology and Environmental Sciences
  - Epidemiology
  - Physics

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
    - Agricultural Sciences
    - Astronomy
    - Bioinformatics
    - Biosurveillance
    - Ecology and Environmental Sciences
    - Epidemiology
    - Physics
    - Reliability and Quality Control

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
  - Agricultural Sciences
  - Astronomy
  - Bioinformatics
  - Biosurveillance
  - Ecology and Environmental Sciences
  - Epidemiology
  - Physics
  - Reliability and Quality Control
  - Signal Detection

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
    - Agricultural Sciences
    - Astronomy
    - Bioinformatics
    - Biosurveillance
    - Ecology and Environmental Sciences
    - Epidemiology
    - Physics
    - Reliability and Quality Control
    - Signal Detection
    - Social Networks

Introduction: Applications

- Scan statistics are used for detecting clusters of rare events.
- Applications:
  - Agricultural Sciences
  - Astronomy
  - Bioinformatics
  - Biosurveillance
  - Ecology and Environmental Sciences
  - Epidemiology
  - Physics
  - Reliability and Quality Control
  - Signal Detection
  - Social Networks
  - Telecommunication Sciences

References

- Glaz, J. and Balakrishnan, N. (Eds.) (1999). *Recent Advances on Scan Statistics*. Birkhauser Publishers, Boston.

References

- Glaz, J. and Balakrishnan, N. (Eds.) (1999). *Recent Advances on Scan Statistics*. Birkhauser Publishers, Boston.
- Glaz, J., Naus, J., Wallenstein, S. (2001). *Scan Statistics*. Springer, New York.

References

- Glaz, J. and Balakrishnan, N. (Eds.) (1999). *Recent Advances on Scan Statistics*. Birkhauser Publishers, Boston.
- Glaz, J., Naus, J., Wallenstein, S. (2001). *Scan Statistics*. Springer, New York.
- Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*. Wiley, New York.

References

- Glaz, J. and Balakrishnan, N. (Eds.) (1999). *Recent Advances on Scan Statistics*. Birkhauser Publishers, Boston.
- Glaz, J., Naus, J., Wallenstein, S. (2001). *Scan Statistics*. Springer, New York.
- Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*. Wiley, New York.
- Fu, J. C. and Lou, W.Y.W. (2003). *Distribution Theory of Runs and Patterns and its Applications*. World Scientific.

References

- Glaz, J. and Balakrishnan, N. (Eds.) (1999). *Recent Advances on Scan Statistics*. Birkhauser Publishers, Boston.
- Glaz, J., Naus, J., Wallenstein, S. (2001). *Scan Statistics*. Springer, New York.
- Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*. Wiley, New York.
- Fu, J. C. and Lou, W.Y.W. (2003). *Distribution Theory of Runs and Patterns and its Applications*. World Scientific.
- Glaz, J, Pozdnyakov, V. and Wallenstein, S. (Eds.) (2009). *Scan Statistics-Methods and Applications*. Springer, Boston.

References

- Glaz, J. and Balakrishnan, N. (Eds.) (1999). *Recent Advances on Scan Statistics*. Birkhauser Publishers, Boston.
- Glaz, J., Naus, J., Wallenstein, S. (2001). *Scan Statistics*. Springer, New York.
- Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*. Wiley, New York.
- Fu, J. C. and Lou, W.Y.W. (2003). *Distribution Theory of Runs and Patterns and its Applications*. World Scientific.
- Glaz, J, Pozdnyakov, V. and Wallenstein, S. (Eds.) (2009). *Scan Statistics-Methods and Applications*. Springer, Boston.
- Duczmal. L. and Buckeridge, D. L. (2006). A workflow scan statistic. *Statistics in Medicine* **25**, 743-754.

References

- Glaz, J. and Balakrishnan, N. (Eds.) (1999). *Recent Advances on Scan Statistics*. Birkhauser Publishers, Boston.
- Glaz, J., Naus, J., Wallenstein, S. (2001). *Scan Statistics*. Springer, New York.
- Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*. Wiley, New York.
- Fu, J. C. and Lou, W.Y.W. (2003). *Distribution Theory of Runs and Patterns and its Applications*. World Scientific.
- Glaz, J, Pozdnyakov, V. and Wallenstein, S. (Eds.) (2009). *Scan Statistics-Methods and Applications*. Springer, Boston.
- Duczmal. L. and Buckeridge, D. L. (2006). A workflow scan statistic. *Statistics in Medicine* **25**, 743-754.
- Glaz, J. and Zhang, Z. (2004). Multiple window discrete scan statistics. *Journal of Applied Statistics* **31**, 967-980.

References

- Glaz, J. and Balakrishnan, N. (Eds.) (1999). *Recent Advances on Scan Statistics*. Birkhauser Publishers, Boston.
- Glaz, J., Naus, J., Wallenstein, S. (2001). *Scan Statistics*. Springer, New York.
- Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*. Wiley, New York.
- Fu, J. C. and Lou, W.Y.W. (2003). *Distribution Theory of Runs and Patterns and its Applications*. World Scientific.
- Glaz, J, Pozdnyakov, V. and Wallenstein, S. (Eds.) (2009). *Scan Statistics-Methods and Applications*. Springer, Boston.
- Duczmal. L. and Buckeridge, D. L. (2006). A workflow scan statistic. *Statistics in Medicine* **25**, 743-754.
- Glaz, J. and Zhang, Z. (2004). Multiple window discrete scan statistics. *Journal of Applied Statistics* **31**, 967-980.
- Glaz, J. and Zhang, Z. (2006). Maximum scan score-type statistic. *Statistics & Probability Lett.* **76**, 1316-1322.

More References

- Glaz, J., Kulldorff, M., Pozdnyakov, V. and Steele, J. M. (2006). Gambling teams and waiting times for patterns in two-state Markov chains. *J. Applied Probability* **43**, 127-140.

More References

- Glaz, J., Kulldorff, M., Pozdnyakov, V. and Steele, J. M. (2006). Gambling teams and waiting times for patterns in two-state Markov chains. *J. Applied Probability* **43**, 127-140.
- Glaz, J., Naus, J. and Wang, X. (2012). Approximations and inequalities for moving sums. *Methodology and Computing in Applied Probability* **14**, 597-616.

More References

- Glaz, J., Kulldorff, M., Pozdnyakov, V. and Steele, J. M. (2006). Gambling teams and waiting times for patterns in two-state Markov chains. *J. Applied Probability* **43**, 127-140.
- Glaz, J., Naus, J. and Wang, X. (2012). Approximations and inequalities for moving sums. *Methodology and Computing in Applied Probability* **14**, 597-616.
- Kulldorff, M. et. al. (2005). A space-time permutation scan statistic for disease outbreak detection. PLOS Medicine, **2**, 1-9.

More References

- Glaz, J., Kulldorff, M., Pozdnyakov, V. and Steele, J. M. (2006). Gambling teams and waiting times for patterns in two-state Markov chains. *J. Applied Probability* **43**, 127-140.
- Glaz, J., Naus, J. and Wang, X. (2012). Approximations and inequalities for moving sums. *Methodology and Computing in Applied Probability* **14**, 597-616.
- Kulldorff, M. et. al. (2005). A space-time permutation scan statistic for disease outbreak detection. PLOS Medicine, **2**, 1-9.
- Neill, D. B., Moore, A. W. and Cooper, G. F. (2006). A Bayesian spatial scan statistic. In Y. Weiss, et. al. ,eds. *Advances in Neural Information Processing Systems* **18**, 1003-1010.

More References

- Glaz, J., Kulldorff, M., Pozdnyakov, V. and Steele, J. M. (2006). Gambling teams and waiting times for patterns in two-state Markov chains. *J. Applied Probability* **43**, 127-140.
- Glaz, J., Naus, J. and Wang, X. (2012). Approximations and inequalities for moving sums. *Methodology and Computing in Applied Probability* **14**, 597-616.
- Kulldorff, M. et. al. (2005). A space-time permutation scan statistic for disease outbreak detection. PLOS Medicine, **2**, 1-9.
- Neill, D. B., Moore, A. W. and Cooper, G. F. (2006). A Bayesian spatial scan statistic. In Y. Weiss, et. al. ,eds. *Advances in Neural Information Processing Systems* **18**, 1003-1010.
- Patil, G. P. et. al. (2004). Detection and delineation of critical areas using echelons and spatial scan statistics with synoptic cellular data. *Environmental and Ecological Statistics* **11**, 139-164.

More References

- Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191-213.

More References

- Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191-213.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2004). False discovery control for random fields. *Journal of the American Statistical Association*, **99**, 1002-1014.

More References

- Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191-213.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2004). False discovery control for random fields. *Journal of the American Statistical Association*, **99**, 1002-1014.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2007). Scan clustering: A false discovery approach. *Journal of Multivariate Analysis*, **98**, 1441-1469.

More References

- Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191-213.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2004). False discovery control for random fields. *Journal of the American Statistical Association*, **99**, 1002-1014.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2007). Scan clustering: A false discovery approach. *Journal of Multivariate Analysis*, **98**, 1441-1469.
- Zhang, Z. and Glaz, J. (2008). Bayesian variable window scan statistics. *Journal of Statistical Planning and Inference*, **138**, 3561-67.

More References

- Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191-213.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2004). False discovery control for random fields. *Journal of the American Statistical Association*, **99**, 1002-1014.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2007). Scan clustering: A false discovery approach. *Journal of Multivariate Analysis*, **98**, 1441-1469.
- Zhang, Z. and Glaz, J. (2008). Bayesian variable window scan statistics. *Journal of Statistical Planning and Inference*, **138**, 3561-67.
- Guerriero, M., Willett, P. and Glaz, J. (2009). Distributed target detection in a sensor network using scan statistics. *IEEE Trans. Signal Processing*, **57**, 2629-2639.

More References

- Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191-213.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2004). False discovery control for random fields. *Journal of the American Statistical Association*, **99**, 1002-1014.
- Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2007). Scan clustering: A false discovery approach. *Journal of Multivariate Analysis*, **98**, 1441-1469.
- Zhang, Z. and Glaz, J. (2008). Bayesian variable window scan statistics. *Journal of Statistical Planning and Inference*, **138**, 3561-67.
- Guerriero, M., Willett, P. and Glaz, J. (2009). Distributed target detection in a sensor network using scan statistics. *IEEE Trans. Signal Processing*, **57**, 2629-2639.
- Glaz, J., Guerriero, M. and Sen, R. (2010). Approximations for the distribution of a scan statistic for 0-1 iid Bernoulli trials in a three dimensional cube. *Method. Comput. Appl. Probab.* **12**, 731-748.

Theory and Methods

- **Bonferroni-type Approximations and Inequalities**

1. Chen, J., Glaz, J., Naus, J. and Wallenstein, S. (2001). Bonferroni-type inequalities for conditional scan statistics. Statistics & Probability Letters 53, 67-77.
2. Chen, J. and Glaz, J. (2002). Approximations for conditional two-dimensional scan statistics. Statistics & Probability Letters 58, 287-296.

1. Wu, T.- L., Glaz, J. and Fu J. C. (2013). Discrete, Continuous and Conditional Variable Window Scan Statistics. *Journal of Applied Probability*, in press.
2. Balakrishnan, N. and Koutras, M. V. (2002). Runs and Scans with Applications. Wiley, New York.
3. Fu, J. C. and Lou, W.Y.W. (2003). Distribution Theory of Runs and Patterns and its Applications. World Scientific, Singapore.

- **Bonferroni-type Approximations and Inequalities**

1. Chen, J., Glaz, J., Naus, J. and Wallenstein, S. (2001). Bonferroni-type inequalities for conditional scan statistics. Statistics & Probability Letters 53, 67-77.
2. Chen, J. and Glaz, J. (2002). Approximations for conditional two-dimensional scan statistics. Statistics & Probability Letters 58, 287-296.

- **Finite Markov Chain Embedding**

1. Wu, T.- L., Glaz, J. and Fu J. C. (2013). Discrete, Continuous and Conditional Variable Window Scan Statistics. *Journal of Applied Probability*, in press.
2. Balakrishnan, N. and Koutras, M. V. (2002). Runs and Scans with Applications. Wiley, New York.
3. Fu, J. C. and Lou, W.Y.W. (2003). Distribution Theory of Runs and Patterns and its Applications. World Scientific, Singapore.

- **Large Deviation Approximations**

1. Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191-213.
2. Loader, C. (1991). Large deviation approximations to distribution of scan statistics. *Advances of Applied Probability* **23**, 751-771.

1. Pozdnyakov, V., Glaz, J., Kulldorff, M. & Steele, M. (2005). A Martingale approach to scan statistics. *Annals Institute of Statistical Mathematics* **57**, 21-37.
2. Glaz, J., Kulldorff, M., Pozdnyakov, V.& Steele, M. (2005). Gambling teams and waiting times in two-state Markov chains. *J. Appl. Probab.*, **43**, no. 1, 127-140.

Theory and Methods

- **Large Deviation Approximations**

1. Siegmund, D. and Yakir, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191-213.
2. Loader, C. (1991). Large deviation approximations to distribution of scan statistics. *Advances of Applied Probability* **23**, 751-771.

- **Martingale Formulation via Betting Systems**

1. Pozdnyakov, V., Glaz, J., Kulldorff, M. & Steele, M. (2005). A Martingale approach to scan statistics. *Annals Institute of Statistical Mathematics* **57**, 21-37.
2. Glaz, J., Kulldorff, M., Pozdnyakov, V.& Steele, M. (2005). Gambling teams and waiting times in two-state Markov chains. *J. Appl. Probab.*, **43**, no. 1, 127-140.

Theory and Methods

- **Perfect Simulation Algorithms**

1. Haiman, G. (2000). Estimating the distribution of scan statistics with high precision. *Extremes* **3**, 349-361.
2. Haiman, G. and Preda, C. (2002). A New method for estimating the distribution of scan statistics for a two-dimensional Poisson process. *Methodology and Computing in Applied Probability* **4**, 393-407.

1. Glaz, J., Naus, J., Roos, M. and Wallenstein, S. (1994). Poisson approximations for the distribution and moments of ordered m-spacings. *Journal of Applied Probability* **31A**, 271-281.
2. Barbour, A. D., Chryssaphinou, O., and Roos, M. (1996). Compound Poisson approximations in systems reliability. *Naval Research Logistics* **43**, 251-264.

- **Perfect Simulation Algorithms**

1. Haiman, G. (2000). Estimating the distribution of scan statistics with high precision. *Extremes* **3**, 349-361.
2. Haiman, G. and Preda, C. (2002). A New method for estimating the distribution of scan statistics for a two-dimensional Poisson process. *Methodology and Computing in Applied Probability* **4**, 393-407.

- **Poisson and Compound-Poisson Approximations**

1. Glaz, J., Naus, J., Roos, M. and Wallenstein, S. (1994). Poisson approximations for the distribution and moments of ordered m-spacings. *Journal of Applied Probability* **31A**, 271-281.
2. Barbour, A. D., Chryssaphinou, O., and Roos, M. (1996). Compound Poisson approximations in systems reliability. *Naval Research Logistics* **43**, 251-264.

- **Product-type Approximations and Inequalities**

1. Glaz, J. and Naus, J. (1991). Tight bounds for scan statistics probabilities for discrete data. *Annals of Applied Statistics* **1**, 306-318.
2. Chen, J. and Glaz, J. (1996). Two-dimensional discrete scan statistics. *Statistics and Probability Letters* **31**, 59-68.

1. Chan, H. P. and Lai, T. L. (2002). Boundary crossing probabilities for scan statistics and their applications to change-point detection. *Methodology and Computing in Applied Probability* **4**, 317-336.
2. Chan, H. P. and Lai, T. L. (2003). Saddle point approximations and nonlinear boundary crossing probabilities for Markov random walks. *Annals of Applied Probability* **13**, 395-429.

- **Product-type Approximations and Inequalities**

1. Glaz, J. and Naus, J. (1991). Tight bounds for scan statistics probabilities for discrete data. *Annals of Applied Statistics* **1**, 306-318.
2. Chen, J. and Glaz, J. (1996). Two-dimensional discrete scan statistics. *Statistics and Probability Letters* **31**, 59-68.

- **Saddle-point Approximations**

1. Chan, H. P. and Lai, T. L. (2002). Boundary crossing probabilities for scan statistics and their applications to change-point detection. *Methodology and Computing in Applied Probability* **4**, 317-336.
2. Chan, H. P. and Lai, T. L. (2003). Saddle point approximations and nonlinear boundary crossing probabilities for Markov random walks. *Annals of Applied Probability* **13**, 395-429.

- **FDR and FDC Methods**

1. Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2004). False discovery control for random fields. *Journal of the American Statistical Association*, **99**, 1002-1014.
2. Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2007). Scan clustering: A false discovery approach. *Journal of Multivariate Analysis*, **98**, 1441-1469.

1. Kulldorff, M. (1999). Spatial scan statistics: Models, calculations and applications. In: Glaz, J. and Balakrishnan, N. (eds). *Scan Statistics and Applications*. Birkhauser, Boston, 303-322.
2. Kulldorff, M. et. al. (2005). A space-time permutation scan statistic for disease outbreak detection. *PLOS Medicine*, **2**, 1-9.

Theory and Methods

- **FDR and FDC Methods**

1. Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2004). False discovery control for random fields. *Journal of the American Statistical Association*, **99**, 1002-1014.
2. Perone-Pacifico, M. Genovese, C. Verdinelli, I. and Wasserman, L. (2007). Scan clustering: A false discovery approach. *Journal of Multivariate Analysis*, **98**, 1441-1469.

- **GLR-type Tests via Simulation**

1. Kulldorff, M. (1999). Spatial scan statistics: Models, calculations and applications. In: Glaz, J. and Balakrishnan, N. (eds). *Scan Statistics and Applications*. Birkhauser, Boston, 303-322.
2. Kulldorff, M. et. al. (2005). A space-time permutation scan statistic for disease outbreak detection. *PLOS Medicine*, **2**, 1-9.

Theory and Methods

- **Linear Programming**
- **Monte Carlo Methods**
- **Order Statistics and Spacings**
- **Symbolic Computing**

Probability Inequalities

- Let $X_1, ..., X_N, ...$ be a sequence of iid normal observations with mean μ and variance σ^2. Let $Y_{r,u} = \sum_{i=r}^{u} X_i$ for $u \geq r \geq 1$. For integers $2 \leq m < N$, where m is the length of the sliding window and N is the specified range of the monitored process, define the *scan statistic*

$$S_{m,N} = \max_{m \leq j \leq N} \left\{ Y_{j-m+1,j} \right\}. \tag{1}$$

Probability Inequalities

- Let $X_1, ..., X_N, \ldots$ be a sequence of iid normal observations with mean μ and variance σ^2. Let $Y_{r,u} = \sum_{i=r}^{u} X_i$ for $u \geq r \geq 1$. For integers $2 \leq m < N$, where m is the length of the sliding window and N is the specified range of the monitored process, define the *scan statistic*

$$S_{m,N} = \max_{m \leq j \leq N} \{Y_{j-m+1,j}\}. \qquad (1)$$

- The sequence $\{Y_{j-m+1,j}; m \leq j \leq N\}$, based on which the scan statistic is defined, contains $N - m + 1$ observations of moving sums of length m.

Probability Inequalities

- Let $X_1, ..., X_N, ...$ be a sequence of iid normal observations with mean μ and variance σ^2. Let $Y_{r,u} = \sum_{i=r}^{u} X_i$ for $u \geq r \geq 1$. For integers $2 \leq m < N$, where m is the length of the sliding window and N is the specified range of the monitored process, define the *scan statistic*

$$S_{m,N} = \max_{m \leq j \leq N} \left\{ Y_{j-m+1,j} \right\}. \quad (1)$$

- The sequence $\left\{ Y_{j-m+1,j}; m \leq j \leq N \right\}$, based on which the scan statistic is defined, contains $N - m + 1$ observations of moving sums of length m.
- The random variables $\left\{ Y_{j-m+1,j}; m \leq j \leq N \right\}$ have a joint multivariate normal distribution with mean vector $(m\mu, ..., m\mu)'$ and variance and covariance matrix $\Sigma = \{\sigma_{i,j}\}$, where $\sigma_{i,i} = m\sigma^2$, $\sigma_{i,j} = 0$, for $|j - i| \geq m$ and $\sigma_{i,j} = (m - k)\sigma^2$, for $|j - i| = k$, $1 \leq k \leq m - 1$.

Probability Inequalities

- Let $X_1, ..., X_N, \ldots$ be a sequence of iid normal observations with mean μ and variance σ^2. Let $Y_{r,u} = \sum_{i=r}^{u} X_i$ for $u \geq r \geq 1$. For integers $2 \leq m < N$, where m is the length of the sliding window and N is the specified range of the monitored process, define the *scan statistic*

$$S_{m,N} = \max_{m \leq j \leq N} \left\{ Y_{j-m+1,j} \right\}. \quad (1)$$

- The sequence $\left\{ Y_{j-m+1,j}; m \leq j \leq N \right\}$, based on which the scan statistic is defined, contains $N - m + 1$ observations of moving sums of length m.
- The random variables $\left\{ Y_{j-m+1,j}; m \leq j \leq N \right\}$ have a joint multivariate normal distribution with mean vector $(m\mu, ..., m\mu)'$ and variance and covariance matrix $\Sigma = \{\sigma_{i,j}\}$, where $\sigma_{i,i} = m\sigma^2$, $\sigma_{i,j} = 0$, for $|j - i| \geq m$ and $\sigma_{i,j} = (m - k)\sigma^2$, for $|j - i| = k$, $1 \leq k \leq m - 1$.
- For $2 \leq m \leq N$ and $-\infty < t < \infty$, let

$$G_{m,t}(N) = P\left(Y_{1,m} < t, Y_{2,m+1} < t, ..., Y_{N-m+1,N} < t\right). \quad (2)$$

Probability Inequalities

- The distribution of the scan statistic $S_{m,N}$ is given by

$$P\left(S_{m,N} < t\right) = G_{m,t}(N). \tag{3}$$

Probability Inequalities

- The distribution of the scan statistic $S_{m,N}$ is given by

$$P\left(S_{m,N} < t\right) = G_{m,t}(N). \tag{3}$$

- The probability that the scan statistic exceeds level t is given by

$$P\left(S_{m,N} \geq t\right) = 1 - G_{m,t}(N). \tag{4}$$

Probability Inequalities

- The distribution of the scan statistic $S_{m,N}$ is given by

$$P\left(S_{m,N} < t\right) = G_{m,t}(N). \tag{3}$$

- The probability that the scan statistic exceeds level t is given by

$$P\left(S_{m,N} \geq t\right) = 1 - G_{m,t}(N). \tag{4}$$

- This scan statistic can be used in detecting a local change in the process mean within a sequence of N observations via testing the null hypothesis of randomness, H_0, that assumes $X_i, 1 \leq i \leq N$, are iid normal random variables with mean μ_0 and variance σ^2.

Probability Inequalities

- The distribution of the scan statistic $S_{m,N}$ is given by

$$P\left(S_{m,N} < t\right) = G_{m,t}(N). \tag{3}$$

- The probability that the scan statistic exceeds level t is given by

$$P\left(S_{m,N} \geq t\right) = 1 - G_{m,t}(N). \tag{4}$$

- This scan statistic can be used in detecting a local change in the process mean within a sequence of N observations via testing the null hypothesis of randomness, H_0, that assumes $X_i, 1 \leq i \leq N$, are iid normal random variables with mean μ_0 and variance σ^2.
- For the alternative hypothesis, H_1, of a local change in μ , one often specifies a segment of m consecutive observations

$$R(i_0, m) = \{i_0, i_0 + 1, ..., i_0 + m - 1\},$$

where $1 \leq i_0 \leq N - m + 1$ is unknown and $2 \leq m \leq N/4$ is the window length. We first discuss the case when m is known.

Probability Inequalities

- Under H_1, for any $i_0 \leq i \leq i_0 + m - 1$, X_i has a normal distribution with mean μ_1 and variance σ^2, where $\mu_1 > \mu_0$. For $i \notin R(i_0, m)$, $X_i's$ are distributed according to the distribution specified by the null hypothesis.

Probability Inequalities

- Under H_1, for any $i_0 \leq i \leq i_0 + m - 1$, X_i has a normal distribution with mean μ_1 and variance σ^2, where $\mu_1 > \mu_0$. For $i \notin R(i_0, m)$, $X_i' s$ are distributed according to the distribution specified by the null hypothesis.
- Let $X_1, ..., X_N, ...$ be iid continuous random variables with mean μ and variance σ^2. The following inequalities are used in approximating the distribution of the scan statistic:

Probability Inequalities

- Under H_1, for any $i_0 \leq i \leq i_0 + m - 1$, X_i has a normal distribution with mean μ_1 and variance σ^2, where $\mu_1 > \mu_0$. For $i \notin R(i_0, m)$, $X_i's$ are distributed according to the distribution specified by the null hypothesis.
- Let $X_1, ..., X_N, ...$ be iid continuous random variables with mean μ and variance σ^2. The following inequalities are used in approximating the distribution of the scan statistic:
- **Theorem (**Glaz, Naus and Wang 2012**)** *For integers $i, m \geq 2, L \geq 1$,*

$$G(N) \geq \frac{G(im)}{\left[1 + \frac{G(Lm-1)-G(Lm)}{G((L+1)m-1)}\right]^{M-im}}, N \geq (i \vee L)m, \qquad (5)$$

$$G(N) \leq G(im)\{1 - [G((L+1)m - 1) - G((L+1)m)]\}^{N-im},$$

for $N \geq (i \vee (L+1))m$.

Probability Inequalities

- The followig inequalities are used for approximating the distribution of the scan statistic for normal data.

Probability Inequalities

- The followig inequalities are used for approximating the distribution of the scan statistic for normal data.
- For choices of i and L in theorem above and $G(3m-1) \geq G(2m-1)G(2m)$ we get:

$$G(N) \geq \frac{G(2m)}{\left[1 + \frac{G(2m-1)-G(2m)}{G(2m-1)G(2m)}\right]^{N-2m}}, \qquad N \geq 2m, \tag{6}$$

Probability Inequalities

- The followig inequalities are used for approximating the distribution of the scan statistic for normal data.
- For choices of i and L in theorem above and $G(3m-1) \geq G(2m-1)G(2m)$ we get:

$$G(N) \geq \frac{G(2m)}{\left[1 + \frac{G(2m-1)-G(2m)}{G(2m-1)G(2m)}\right]^{N-2m}}, \qquad N \geq 2m, \qquad (6)$$

-

$$G(N) \leq G(2m)\{1 - [G(2m-1) - G(2m)]\}^{N-2m}, \qquad N \geq 2m. \quad (7)$$

Probability Inequalities

- The followig inequalities are used for approximating the distribution of the scan statistic for normal data.
- For choices of i and L in theorem above and $G(3m-1) \geq G(2m-1)G(2m)$ we get:

$$G(N) \geq \frac{G(2m)}{\left[1 + \frac{G(2m-1)-G(2m)}{G(2m-1)G(2m)}\right]^{N-2m}}, \qquad N \geq 2m, \qquad (6)$$

- $$G(N) \leq G(2m)\{1 - [G(2m-1) - G(2m)]\}^{N-2m}, \qquad N \geq 2m. \quad (7)$$

- We expect these bounds to be tight for a large value of t, since they converge as $G(2m) \to 1$ and $G(2m-1) - G(2m) \to 0$, which holds as $t \longrightarrow \infty$.

Probability Inequalities

- The followig inequalities are used for approximating the distribution of the scan statistic for normal data.
- For choices of i and L in theorem above and $G(3m-1) \geq G(2m-1)G(2m)$ we get:

$$G(N) \geq \frac{G(2m)}{\left[1+\frac{G(2m-1)-G(2m)}{G(2m-1)G(2m)}\right]^{N-2m}}, \qquad N \geq 2m, \tag{6}$$

- $$G(N) \leq G(2m)\{1-[G(2m-1)-G(2m)]\}^{N-2m}, \qquad N \geq 2m. \tag{7}$$

- We expect these bounds to be tight for a large value of t, since they converge as $G(2m) \rightarrow 1$ and $G(2m-1) - G(2m) \rightarrow 0$, which holds as $t \longrightarrow \infty$.
- In Glaz, Naus and Wang (2012), inequalities for expected values and variances of a stopping time for moving sums are evaluated via the R algorithms in Genz and Bretz (2009).

- A Markov-type approximation for $G(N)$ based on a method introduced in Naus (1982). Let $N = Km + v$, where $K \geq 3$, $m \geq 2$ and $0 \leq v \leq m-1$ are integers. Then, for $2 \leq L \leq H-1$

$$G(N) = P\left\{\max_{m \leq k \leq N} Y_{k-m+1,k} < t\right\} = P\left(\cap_{j=1}^{K} E_j\right)$$
$$= P\left(\bigcap_{i=1}^{L-1} E_i\right) \prod_{j=L}^{K} P\left(E_j | \cap_{h=1}^{j-1} E_h\right), \quad (8)$$

where for $1 \leq j \leq K-1$

$$E_j = \left(\max_{jm \leq k \leq (j+1)m} Y_{k-m+1,k} < t\right),$$

which can be interpreted as the event of no exceedance of level t within a block of $m+1$ consecutive partial sums of length m, and

$$E_K = \left(\max_{Km \leq k \leq Km+v} Y_{k-m+1,k} < t\right).$$

Product-type approximations

- By conditioning on the most recent past $L \geq 2$ events E_j, in (8) we get the following approximation for $G(M)$:

$$G(N) \approx P\left(\bigcap_{i=1}^{L-1} E_i\right)\left[\prod_{j=L}^{K-1} P\left(E_j | \bigcap_{h=j-L+1}^{j-1} E_h\right)\right] P\left(E_K | \bigcap_{p=K-L+1}^{K-1} E_p\right)$$

$$= P\left(\bigcap_{i=1}^{L} E_i\right)\left\{\prod_{j=L+1}^{K-1}\left[\frac{P\left(\bigcap_{h=j-L+1}^{j} E_h\right)}{P\left(\bigcap_{h=j-L+1}^{j-1} E_h\right)}\right]\right\}\frac{P\left(\bigcap_{p=K-L-1}^{K} E_p\right)}{P\left(\bigcap_{p=K-L-1}^{K-1} E_p\right)}$$

$$= G((L+1)m)\left[\frac{G((L+1)m)}{G(Lm)}\right]^{K-L-1}\frac{G(Lm+v)}{G(Lm)}. \quad (9)$$

- For $N = Km$ and $L = 2$ the above approximation reduces to

$$G(N) \approx G(3m)\left[\frac{G(3m)}{G(2m)}\right]^{K-3}. \quad (10)$$

Product-type approximations:Quasi-stationarity property

- Let $X_1, ..., X_N, ...$ be iid continuous random variables with mean μ and variance σ^2.
- For $m \geq 2, j \geq 1$, let

$$q_j = P\left(Y_{j+1,\, j+m} \leq t \middle| Y_{i,i+m-1} \leq t; 1 \leq i \leq j.\right)$$

- **Theorem** (Glaz and Johnson 1988): If $0 < P(X_1 \leq t/m) < 1$, then $\lim_{j\to\infty} q_j = q$, where $0 < q < 1$.
- The proof of the theorem is based on the R-theory of Markov chains. One can show that for $m = 2$ the $q_j's$ oscilate about q. This property does not extend for $m \geq 3$, even though numerically one observes an oscilatory pattern of convergence of q_j to q.

Haiman approximation

- Haiman (1999 and 2007) derived accurate approximations for $G(M)$ for iid discrete random variables. These approximations are valid as well for iid continuous random variables.
- A nice feature of these approximations is that a sharp error bound can be easily evaluated.
- For the problem at hand, for $N \geq 3m$, the following approximation for $G(N)$ is obtained from Haiman (2007, Corollary 2):

$$G(N) \approx \frac{2G(2m) - G(3m)}{\left[1 + G(2m) - G(3m) + 2\left(G(2m) - G(3m)\right)^2\right]^{N/m}}, \quad (11)$$

- with an error bound of approximately

$$3.3[1 - G(2m)]^2 N/m. \quad (12)$$

A Multiple Window Scan Statistics

- Let $2 \leq m_1 < m_2 < ..... < m_n$ be a given sequence of window lengths associated with scan statistics $S_{m_1}, ..., S_{m_n}$, respectively.

A Multiple Window Scan Statistics

- Let $2 \leq m_1 < m_2 < ..... < m_n$ be a given sequence of window lengths associated with scan statistics $S_{m_1}, ..., S_{m_n}$, respectively.
- Since the size of the rectangular window m is unknown, for testing H_0 vs H_1 we propose the following test statistic:

$$P_{\min} = \min\{p_j; 1 \leq j \leq n\}, \tag{13}$$

the *minimum P-value statistic,* which is based on n fixed window size scan statistics: $S_{m_1}, ..., S_{m_n}$, where $2 \leq m_j < m_{j+1} \leq N - 1$, $1 \leq j \leq n - 1$, and $p_j = P\left(S_m \geq k_j\right)$, is the observed p-value.

A Multiple Window Scan Statistics

- Let $2 \leq m_1 < m_2 < ..... < m_n$ be a given sequence of window lengths associated with scan statistics $S_{m_1}, ..., S_{m_n}$, respectively.
- Since the size of the rectangular window m is unknown, for testing H_0 vs H_1 we propose the following test statistic:

$$P_{\min} = \min\{p_j; 1 \leq j \leq n\}, \quad (13)$$

the *minimum P-value statistic,* which is based on n fixed window size scan statistics: $S_{m_1}, ..., S_{m_n}$, where $2 \leq m_j < m_{j+1} \leq N - 1$, $1 \leq j \leq n - 1$, and $p_j = P\left(S_m \geq k_j\right)$, is the observed p-value.

- A simulation algorithm is used to implement this multiple window scan statistic and evaluate its power.

A Multiple Window Scan Statistic

- For $1 \leq j \leq n$, let t_j be the observed value of S_{m_j} and $p_j = P(S_{m_j} \geq t_j \mid H_0)$ the associated p-value. Since the exact distribution for the P_{min} statistic is unknown, for a given significant level α, the critical value p_α,

$$P_{H_0}\left(P_{min} \leq p_\alpha\right) = \alpha,$$

has to be evaluated via simulation.

A Multiple Window Scan Statistic

- For $1 \leq j \leq n$, let t_j be the observed value of S_{m_j} and $p_j = P(S_{m_j} \geq t_j \mid H_0)$ the associated p-value. Since the exact distribution for the P_{min} statistic is unknown, for a given significant level α, the critical value p_α,

$$P_{H_0}\left(P_{min} \leq p_\alpha\right) = \alpha,$$

  has to be evaluated via simulation.
- In each run of the simulation, we generate N observations under the null hypothesis. Then we scan the whole region with multiple moving windows of sizes m_1, m_2, ... and m_n, and record the observed values of the fixed window scan statistics, $S_{m_1}, \ldots., S_{m_n}$, denoted by $t_1, t_2, \ldots, t_n$, respectively.

A Multiple Window Scan Statistic

- For $1 \leq j \leq n$, let t_j be the observed value of S_{m_j} and $p_j = P(S_{m_j} \geq t_j \mid H_0)$ the associated p-value. Since the exact distribution for the P_{min} statistic is unknown, for a given significant level α, the critical value p_α,

$$P_{H_0}\left(P_{min} \leq p_\alpha\right) = \alpha,$$

  has to be evaluated via simulation.
- In each run of the simulation, we generate N observations under the null hypothesis. Then we scan the whole region with multiple moving windows of sizes m_1, m_2, ... and m_n, and record the observed values of the fixed window scan statistics, $S_{m_1}, ..., S_{m_n}$, denoted by $t_1, t_2, ..., t_n$, respectively.
- Then, a Monte-Carlo R algorithm is employed to evaluate the observed p values: $p_j = P(S_{m_j} \geq t_j \mid H_0)$, $1 \leq j \leq n$.

A Multiple Window Scan Statistic

- For $1 \leq j \leq n$, let t_j be the observed value of S_{m_j} and $p_j = P(S_{m_j} \geq t_j \mid H_0)$ the associated p-value. Since the exact distribution for the P_{min} statistic is unknown, for a given significant level α, the critical value p_α,

$$P_{H_0}\left(P_{min} \leq p_\alpha\right) = \alpha,$$

  has to be evaluated via simulation.
- In each run of the simulation, we generate N observations under the null hypothesis. Then we scan the whole region with multiple moving windows of sizes m_1, m_2, ... and m_n, and record the observed values of the fixed window scan statistics, $S_{m_1}, \ldots, S_{m_n}$, denoted by $t_1, t_2, \ldots, t_n$, respectively.
- Then, a Monte-Carlo R algorithm is employed to evaluate the observed p values: $p_j = P(S_{m_j} \geq t_j \mid H_0)$, $1 \leq j \leq n$.
- The minimum of value of these p values is recorded and this process is repeated $10,000$ times. Based on that, an approximate $\alpha * 100$ percentile of the distribution of $P_{min}^{(1)}$ statistic is obtained.

Numerical Results

- Inequalities and approximations for a fixed window scan statistic for normal data with $\mu = 0$ and $\sigma^2 = 1$, $N = 1000$, $m = 50$:

t	20	23	24	25	26	27	28
LB	.2530	.0729	.0479	.0388	.0221	.0121	.0074
$A\ 1$	.2601	.0851	.0551	.0350	.0216	.0130	.0077
$A\ 2$	.2587	.0847	.0551	.0349	.0214	.0132	.0077
EB	.	1.84−3	7.39−4	2.84−4	1.05−4	3.63−5	1.21−5
UB	.2607	.0886	.0613	.0390	.0252	.0121	.0085

Numerical Results

- Power study to evaluate the performance of the multiple window scan statistic: for normal data, $H_0 : \mu = 0$, $\sigma^2 = 1$, $N = 250$
- $\alpha =$ Pr Type I Error, $\mu_1 =$ the mean under the aternative in a subsequence of n observations.

n	μ_1	$P_{\min}$	S_5	S_{10}	S_{15}	S_{20}	S_{25}
10	.5	.091	.074	.100	.076	.061	.067
	1	.424	.339	.468	.317	.228	.189
	1.5	.909	.828	.926	.770	.589	.465
15	.5	.150	.116	.149	.155	.119	.099
	1	.742	.506	.695	.773	.618	.505
	1.5	.993	.932	.987	.994	.971	.921
20	.5	.247	.153	.209	.263	.264	.189
	1	.895	.602	.819	.878	.913	.834
	1.5	1.0	.981	.997	1.0	1.0	1.0
α		.05	.045	.048	.047	.046	.063

Scan Statistics for Time Series Data

- Let $X_1, ..., X_M$ be a sequence of observations from an AR(1) process, $X_t = \theta X_{t-1} + \omega_t$, where ω_t is a Gaussian white noise with mean $\mu = 0$ and variance $\sigma^2 = 1$. Since $X_t's$ follow a multivariate normal distribution, $\{Y_{i-m+1,i}; m \leq i \leq M\}$ have a multivariate normal distribution with zero mean vector and covariance matrix $\Sigma = \{\sigma_{i,j}\}$, where $\sigma_{i,j} = cov(Y_{i,i+m-1}, Y_{j,j+m-1})$.
- A routine derivation, yields the following covariance matrix:

$$\sigma_{i,j} = \begin{cases} \frac{\theta}{(1-\theta)^4}(1-\theta^{j+m-i})(1-\theta^{i-j}) + \frac{\theta^{j+m-i+1}}{(1-\theta)^4}(1-\theta^{i-j})^2 \\ +\frac{j+m-i}{(1-\theta)^2} + \frac{2\theta}{(1-\theta)^3}[j+m-1-i-\frac{\theta}{1-\theta}(1-\theta^{j+m-1-i})] \\ +\frac{\theta}{(1-\theta)^4}(1-\theta^{i-j})(1-\theta^{j+m-i}), i-j<m \\ \frac{1}{1-\theta^2}\{m+\frac{2\theta}{1-\theta}[m-1-\frac{\theta}{1-\theta}(1-\theta^{m-1})]\}, i=j \\ \theta^{i-j-m+1}\frac{(1-\theta^m)^2}{(1-\theta)^2}, \text{ otherwise.} \end{cases}$$

Scan Statistics for Time Series Data

- Given the mean vector and covariance matrix, we can utilize the R algorithms by Genz and Bretz (2009) to approximate the distribution $G(M)$ for a fixed window scan statistic and the multiple window scan statistic P_{min}.
- For an AR(2) model, $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \omega_t$, where ω_t is the Gaussian white noise with mean $\mu = 0$ and variance $\sigma^2 = 1$,the $X_t' s$ follow a multivariate normal distribution with the following ACF:

$$\gamma_h = \begin{cases} \frac{1-\theta_2}{1-\theta_2-\theta_1^2-\theta_2\theta_1^2-\theta_2^2+\theta_2^3}, & \text{when} \quad \text{if } h=0, \\ \gamma_0 \frac{\theta_1}{1-\theta_2}, & \text{when} \quad \text{if } h=1, \\ \gamma_0[\theta_1\gamma_{h-1} + \theta_2\gamma_{h-2}], & \text{when} \quad \text{if } h>1. \end{cases}$$

- Then $\{Y_{i-m+1,i}; m \leq i \leq M\}$ a multivariate normal distribution with a mean vector of zeros and covariance matrix $\Sigma = \{\sigma_{i,j}\}$, which can be derived similarly as in the AR(1) process. The explicit form of the covariance matrix is omitted here for simplicity. Wang and Glaz (2013) investigated the performance of multiple window scan

Numerical Results

- Power study to evaluate the performance of the multiple window scan statistic $P_{\min}$: for AR(1) data, $\theta = .1$, $N = 1500$
- α = Pr Type I Error, μ_1 = the mean under the aternative in a subsequence of n observations in the white noise component.

n	μ_1	$P_{\min}$	S_5	S_{10}	S_{15}	S_{20}	S_{25}
10	.5	.076	.075	.063	.056	.062	.05
	1	.292	.211	.337	.172	.124	.101
	1.5	.797	.649	.841	.554	.388	.285
15	.5	.103	.072	.087	.093	.085	.074
	1	.532	.273	.494	.594	.407	.297
	1.5	.973	.810	.960	.989	.915	.800
20	.5	.115	.068	.091	.113	.120	.100
	1	.762	.378	.615	.759	.818	.683
	1.5	1.0	.905	.992	1.0	1.0	.992
α		.050	.037	.052	.052	.051	.052

Series D Data Set from Box and Jenkins (1978)

- This data set consists of 310 hourly uncontrolled viscosity readings of a chemical process. This data set has been modeled via an AR(1) process in Box and Jenkins (1978), with estimated parameters: $\theta = 0.87$, and $\sigma^2 = 0.09$

Series D Data Set from Box and Jenkins (1978)

- This data set consists of 310 hourly uncontrolled viscosity readings of a chemical process. This data set has been modeled via an AR(1) process in Box and Jenkins (1978), with estimated parameters: $\theta = 0.87$, and $\sigma^2 = 0.09$
- To evaluate the performance of the multiple window scan statistic, we introduced a change in the Gaussian white noise component at a random location. We employed a similar algorithm to the one outlined above to perform a power study that is presented in the table below. A simulation with 10,000 trial has been used to simulate the power.

Series D Data Set from Box and Jenkins (1978)

- This data set consists of 310 hourly uncontrolled viscosity readings of a chemical process. This data set has been modeled via an AR(1) process in Box and Jenkins (1978), with estimated parameters: $\theta = 0.87$, and $\sigma^2 = 0.09$
- To evaluate the performance of the multiple window scan statistic, we introduced a change in the Gaussian white noise component at a random location. We employed a similar algorithm to the one outlined above to perform a power study that is presented in the table below. A simulation with 10,000 trial has been used to simulate the power.
- The multiple window scan statistic outperformed the fixed window scan statistics, with an incorrectly specified window size where a change in mean has occurred. A discrepancy in some of the results could have resulted from the model lack of fit.

Numerical Results

- Series D data set from Box and Jenkins (1978).

- Series D data set from Box and Jenkins (1978).
- α = Pr Type I Error, μ_1 = the mean under the aternative in a subsequence of n observations in the white noise component.

n	μ_1	$P_{\min}$	S_5	S_{10}	S_{15}	S_{20}	S_{25}
10	.15	.142	.170	.260	.035	0	0
	.20	.584	.588	.628	.330	.100	0
	.25	.703	.731	.710	.641	.403	.262
15	.15	.394	.234	.379	.473	.226	.161
	.20	.704	.725	.700	.664	.622	.520
	.25	.841	.845	.834	.864	.756	.731
20	.15	.625	.324	.499	.574	.617	.500
	.20	.778	.787	.774	.750	.753	.737
	.25	.932	.889	.914	.910	.942	.805
α		.051	.040	.054	.049	.059	.040

.

Summary and Future Work

- Two dimensional continuous-type data sets
- Scan statistics for graphs
- Non-homogeneous processes
- Three dimensional scan statistics
- Conditional-type scan statistics