

A MARTINGALE APPROACH TO SCAN STATISTICS

VLADIMIR POZDNYAKOV¹, JOSEPH GLAZ¹, MARTIN KULLDORFF² AND J. MICHAEL STEELE³

¹*Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269-4120, U.S.A.*

²*Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 133 Brookline Avenue, Boston, MA 02215-3920, U.S.A.*

³*Wharton School, Department of Statistics, University of Pennsylvania, Huntsman Hall 447, Philadelphia, PA 19104, U.S.A.*

(Received October 23, 2003; revised March 12, 2004)

Abstract. Scan statistics are commonly used in biology, medicine, engineering and other fields where interest is in the probability of observing clusters of events in a window at an unknown location. Due to the dependent nature of the number of events in a large number of overlapping window locations, even approximate solutions for the simplest scan statistics may require elaborate calculations. We propose a new martingale method which allows one to approximate the distribution for a wide variety of scan statistics, including some for which analytical results are computationally infeasible.

Key words and phrases: Scan, run, pattern, martingale, stopping time.