

# Spatial CUSUM Chart Based Method for Rapid Detection of Outbreaks

Sesha K. Dassanayaka\*

Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO, USA

## Objective

To develop a computationally simple and fast algorithm for rapid detection of outbreaks producing easily interpretable results.

## Introduction

Since the release of anthrax in October of 2001, there has been increased interest in developing efficient prospective disease surveillance schemes. Poisson CUSUM is a control chart-based method that has been widely used to detect aberrations in disease counts in a single region collected over fixed time intervals. Over the past few years, different methods have been proposed to extend Poisson CUSUM charts to capture the spatial association among several regions simultaneously.

In the proposed method, we extend an algorithm<sup>2</sup> in industrial process control using multiple Poisson CUSUM charts to the spatial setting. The spatial association among regions is captured using the method proposed by Raubertas<sup>3</sup>, which has been successfully applied in several prospective surveillance schemes. Also, to improve the power of the traditional multiple Poisson CUSUM charts, Poisson CUSUM charts were used along with fault discovery rate (FDR) control techniques<sup>2</sup>.

## Methods

We start by defining the overall the false discovery rate which allows the user to define an acceptable false alarm rate. Then for each region  $m$ , we sum up the counts of all the immediate neighbors, creating  $m$  spatially correlated neighborhoods. A Poisson CUSUM statistic is then calculated for each neighborhood. To calculate the Poisson CUSUM statistic, the in-control Poisson parameter needs to be calculated using past data and a tolerable out-of-control parameter needs to be specified by the user. A computationally simple random walk method is used to find p-values for each regional Poisson CUSUM statistic<sup>2</sup> instead of the computationally intensive Markov chain-based method while producing comparable results.

The use of p-values to identify alarms makes the procedure easy to interpret for a wider audience in comparison to the traditional average run length (ARL) and cut-off thresholds that are only used in an industrial process control setting. The use of FDR to control multiple testing allows the utilization of popular multiple testing techniques such as Benjamini-Yekutieli<sup>1</sup> procedure.

Once the p-values are calculated, we use the general version of the Benjamini-Yekutieli procedure for multiple testing of the p-values to identify neighborhoods with unusually high disease counts; currently, we are considering alternative resampling based procedures to gain even more power.

## Results

We developed a grid of neighboring regions and simulated Poisson counts with constant mean during the first half of the time period. During the second half, we introduced a step-increase in mean at varying degrees of intensity to simulate an outbreak in neighboring regions. Then we ran the simulation to calculate an independent Poisson CUSUM statistic for each region. Fixing the fault discovery

rate at 0.05, we used the Benjamin-Hochberg<sup>1</sup> procedure to determine the speed at which the procedure identifies the simulated outbreak.

Next, for the same simulated data, we cumulated the Poisson counts of the neighboring regions and used Benjamini-Yekutieli procedure to detect alarms from the spatially correlated neighborhood statistics. The results provide convincing evidence that using the neighborhoods - instead of independent regions - significantly reduces the time for detection while increasing the correct identification of time periods of the simulated outbreak.

The method proposed by Raubertas acknowledges spatial association but does not account for it, as it simply cumulates neighboring counts without using a measure of spatial correlation to calculate neighborhood statistics. To further improve, we are currently investigating spatial statistical methods to account for spatial correlation.

Finally, we plan to apply this procedure to lower respiratory infection data during 1996-1999 from multiple Boston clinics.

## Conclusions

We have developed a simple, efficient algorithm for prospective disease surveillance that is relatively easy to interpret without a high level of statistical expertise. Our preliminary results from simulation studies provide evidence of the strengths of the method proposed.

## Keywords

spatial CUSUM charts; biosurveillance; disease surveillance; multiple control charts; false discovery rate

## References

1. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 2001; 29(4):1165-1188.
2. Li Y, Tsung F. Multiple attribute control charts with false discovery rate control. *Qual. Reliab. Eng. Int.* 2012; 28(8):857-871.
3. Raubertas, RF. An analysis of disease surveillance data that uses the geographical locations of the reporting units. *Statist. Med.* 1989; 8:267-271.

\*Sesha K. Dassanayaka

E-mail: sesha.dassanayaka@ucdenver.edu

