Sufficient reduction methods for multivariate surveillance

Sawaporn Siripanthana* and Eleanor Stillman

School of Mathematics and Statistics, The University of Sheffield, Sheffield, United Kingdom

Objective

To reduce the dimensionality of p-dimensional multivariate series to a univariate series by deriving sufficient statistics, which take into account all the information in the original data, correlation within series (CWS) and correlation between series (CBS).

Introduction

Parallel surveillance, separate monitoring of each continuous series, has been widely used for multivariate surveillance; however, it has severe limitations. First, it faces the problem of multiplicity from multiple testing. Also, the ignorance of CBS reduces the performance of outbreak detection if data are truly correlated. Finally, since health data are normally dependent over time, CWS is another issue that should be taken into account. Sufficient reduction methods are used to reduce the dimensionality of a simple multivariate series to a univariate series, which has been proved to be sufficient for monitoring a mean shift in multivariate surveillance (1, 2). Having considered the sufficiency property and the nature of health data, we propose a sufficient reduction method for detecting a mean shift in multivariate series where CWS and CBS are taken into account.

Methods

Wessman (1) and Frisén et al (2) proposed a sufficient reduction method used for monitoring a mean shift in multivariate series where observations are assumed independent. Also, the former allows for CBS while the latter does not. In this study, we further develop sufficient reduction methods by taking CWS and CBS into account. At each time point, data from p-dimensional multivariate series are used to calculate sufficient statistics derived from the likelihood ratio between out of control and in control states. The evaluation of this method is by simulation study, where bivariate series are generated daily from different sets of parameters (whether or not CWS and/or CBS are present) (3). Detection of a mean shift, which is 2, 3 or 4 times standard deviation of background data, is investigated. A EWMA chart is used to monitor the resultant series of sufficient statistics, and the conditional expected delays (CED) and false alarm rates (FAR) (4) from four methods are compared.

Results

Three examples from our simulation study are shown in Table 1. Data are generated from three different sets of parameters (CWS (φ) and CBS (ρ)), and the aim is to detect a shift in mean of 2s.d. Dataset 1 has no CWS and CBS ($\varphi = 0$ and $\rho = 0$). Dataset 2 includes CWS ($\varphi = 0.4$), while dataset 3 presents both CWS and CBS ($\varphi = 0.4$ and $\rho = 0.3$). For all datasets, the parallel method gives longer delays compared with other methods. In the case of dataset 1 (no CWS and CBS), the last three methods perform similarly. When CWS is present (dataset 2), the proposed method performs better than the others with slightly lower delay and much lower FAR. This pattern is repeated when CBS is incorporated (dataset 3).

Conclusions

When CWS is present, CWS should be taken into account to the sufficient reduction method as ignoring CWS delays detection and gives more false alarms. Sufficient reduction methods derived for independent observations (1, 2) do not take CWS into account; therefore, the effect of CWS is still present in their derived series of sufficient statistics. This effect violates the assumptions of EWMA chart, for which data are assumed to be independent and then produces a high FAR. Although the sufficient reduction method proposed by Wessman (1) allows for CBS, it does not allow CWS. Incorporating both CWS and CBS in our proposed sufficient reduction method substantially improves the performance in detecting a mean shift in multivariate surveillance data.

Keywords

Sufficient statistics; multivariate surveillance; autocorrelation

References

1. Wessman P. Some principles for surveillance adopted for multivariate process with a common change point. Commun Stat Theor Meth. 1998;27:1143–61.
2. Frisén M, Andersson E, Schiöler L. Sufficient reduction in multivariate surveillance. Research report, Statistical Research Unit, University of Gothenburg; 2009.
3. Lotze T, Shmueli G, Yahav I. Simulating multivariate syndromic time series and outbreak signatures. Robert H. Smith School of Business, University of Maryland; 2007.
4. Frisén M, Andersson E, Schiöler L. Evaluation of multivariate surveillance. J Appl Stat. 2010;37:2089–100.

*Sawaporn Siripanthana
E-mail: smp08ss@sheffield.ac.uk

Emerging Health Threats Journal 2011. © 2011 S. Siripanthana and E. Stillman This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Emerging Health Threats Journal 2011, **4**: 10942 - DOI: 10.3402/ehtj.v4i0.10942
(page number not for citation purpose)