Some Theoretical Properties of the Multivariate Exponentially Weighted Moving Covariance Matrix

Edgard M. Maboudou

Department of Statistics, U of Central Florida, Fl

July 29, 2009

Introduction and background

- A process is in statistical control when it is in presence of common cause (unavoidable cause) variation, which is intrinsic to the process and will always be present
- A process is out of statistical control when it is in presence of special cause variation, which stems from external sources
- Statistical process control (SPC) procedures can help you monitor process behavior.
- Arguably the most successful SPC tool is the control chart, originally developed by Walter Shewhart in the early 1920s.

Statistical Process Control (SPC) Charts

- SPC detects, diagnoses changes from in-control to out-of-control situation.
- Traditionally important in manufacturing; now also used in health care, government, education, business,...
- Some applications: fraud detection, anomalies detection, regime switch, outbreak detection, bio-surveillance, cyber surveillance
- Decide whether chart support ‘still in control’, or ‘has gone out of control’

- If process still in control, leave process alone
- If out of control, diagnose, identify problems, fix, and get going again
- Examples of univariate control charts: $\bar{X}$ chart, EWMA chart, CUSUM chart.
- Examples of multivariate control charts: Hotelling T^2 chart, MEWMA chart, MCUSUM chart.

Litterature Review

- Several methodologies for monitoring the covariance matrix
- Generalized variance of Montgomery and Wadsworth (1972), charting the determinant of $\mathbf{S}$
- Charting the trace of the covariance matrix $\mathbf{S}$, Reynolds and Cho (2006)
- Generalized likelihood ratio (GLR) statistic, Alt (1984)

Problems with previous methods

- The previous methods are Shewhart type charts
- These charts are effective for large transient changes in the covariance matrix
- However, they are less suitable for small persistent shifts

Definition of the MEWMC

- Suppose in control data vectors $\mathbf{X}_n$ normal, iid, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are assumed known)
- Let λ such that $0 \leq \lambda \leq 1$.
- Multistandardize the process readings $\mathbf{X}_n$– Transform the $\mathbf{X}_n$ to $\mathbf{U}_n$ such that the $\mathbf{U}_n$'s are N($\mathbf{0}$,$\mathbf{I}_p$)
- Define the sequence $\mathbf{S}_n$ by the recursion $\mathbf{S}_0 = \mathbf{I}_p$, and for $n = 1, 2, ...$

$$\mathbf{S}_n = (1 - \lambda)\mathbf{S}_{n-1} + \lambda \mathbf{U}_n \mathbf{U}_n' \tag{3.1}$$

Some properties of the MEWMC

While the process is in-control,

- By decomposing $\mathbf{S}_{n-1}$ in terms of $\mathbf{S}_{n-2}$, $\mathbf{S}_{n-2}$ in terms of $\mathbf{S}_{n-3}$, and so on, it follows

Lemma

$$\mathbf{S}_n = (1-\lambda)^n \mathbf{I}_p + \lambda \sum_{k=0}^{n-1} (1-\lambda)^k \mathbf{U}_{n-k} \mathbf{U}'_{n-k} \qquad (3.2)$$

- The above formula shows that $\mathbf{S}_n$ is a linear combination of the identity matrix weighted by a coefficient $(1-\lambda)^n$ and the random matrices $\mathbf{U}_1\mathbf{U}_1', \ldots, \mathbf{U}_n\mathbf{U}_n'$ weighted by the coefficients $\lambda(1-\lambda)^{n-1}, \ldots, \lambda$
- For this reason, the sequence $\mathbf{S}_1, \ldots \mathbf{S}_n, \ldots$ is called a Multivariate Exponentially Weighted Moving Covariance Matrix (MEWMC) sequence.
- when $\lambda \rightarrow 0$, the sequence $\mathbf{S}_1, \ldots \mathbf{S}_n$ tends to be a smoother version of the initial sequence $\mathbf{U}_1\mathbf{U}_1', \ldots, \mathbf{U}_n\mathbf{U}_n'$
- when $\lambda = 0$, then $\mathbf{S}_n = \mathbf{S}_{n-1} = \ldots = \mathbf{S}_0 = \mathbf{I}_p$

- when $\lambda \to 1$, the sequence $\mathbf{S}_1, \ldots \mathbf{S}_n$ tends to be a copy of the initial sequence $\mathbf{U}_1\mathbf{U}_1', \ldots, \mathbf{U}_n\mathbf{U}_n'$
- when $\lambda = 1$, then $\mathbf{S}_n = \mathbf{U}_n\mathbf{U}_n'$ for $n \geq 1$
- I will assume that $0 < \lambda < 1$.
- As $\mathbf{S}_n$ is a covariance matrix like, it must possess properties of a covariance matrix. So, it follows

Theorem

$\mathbf{S}_n$ *is a positive definite matrix.*

Independence of the $\mathbf{S}_n$

- The random variables $\mathbf{U}_1, \ldots, \mathbf{U}_n$ are, by definition, independent
- So are the random matrices $\mathbf{U}_1\mathbf{U}_1', \ldots, \mathbf{U}_n\mathbf{U}_n'$ as a consequence
- However, the random matrices $\mathbf{S}_1, \ldots \mathbf{S}_n$ are not independent

Theorem

The sequence $\mathbf{S}_1, \ldots \mathbf{S}_n, \ldots$ has a Markov property.

- The sequence $\mathbf{S}_1, \ldots \mathbf{S}_n, \ldots$ is a Markov chain (nonstationary)

Expectations of the MEWMC

- By using lemma 3.1 , we can show using an argument by induction that

Lemma

$$E(\mathbf{S}_n) = \mathbf{I}_p \tag{3.3}$$

Lemma

$$Cov(\mathbf{S}_n) = \frac{2\lambda}{2-\lambda}\left[1-(1-\lambda)^{2n}\right](\mathbf{I}_p \otimes \mathbf{I}_p)\,. \qquad (3.4)$$

- As $n \to \infty$,

$$Cov(\mathbf{S}_n) = \frac{2\lambda}{2-\lambda}(\mathbf{I}_p \otimes \mathbf{I}_p)$$

- So, the asymptotic covariance of the process $\mathbf{S}_n$ is $\frac{2\lambda}{2-\lambda}(\mathbf{I}_p \otimes \mathbf{I}_p)$

Trace of the MEWMC

- Trace of $\mathbf{S}_n$

Lemma

$$tr(\mathbf{S}_n) = (1-\lambda)^n p + \lambda \sum_{i=1}^{n} d_i \left(\sum_{j=1}^{p} \mathbf{U}_{ij}^2 \right) \tag{3.5}$$

where tr is the trace operator.

- By using lemma 3.4, it is straightforward to show that

Lemma

$$E[tr(\mathbf{S}_n)] = p. \tag{3.6}$$

Lemma

$$V(tr(\mathbf{S}_n)) = 2p\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2n}]. \tag{3.7}$$

- The asymptotic variance of the trace of the process $\mathbf{S}_n$ is $2p\frac{\lambda}{2-\lambda}$

Comments

- As the random matrices $\mathbf{S}_1, \ldots \mathbf{S}_n$ are not independent, it is very hard to derive the distribution of the sequence $\mathbf{S}_n$
- Also, want the distribution of the trace of $\mathbf{S}_n$
- want the generalized sample variance of $\mathbf{S}_n$ as well as its distribution

The MEWMC statistic

- Define MEWMC statistic to compare the matrix $\mathbf{S}_n$ with the identity
- Use the statistic

$$c_n = tr(\mathbf{S_n}) - log|\mathbf{S_n}| - p \tag{4.1}$$

- tr = trace of a matrix, $|\mathbf{S_n}|$ is the determinant of $\mathbf{S_n}$
- Exact distribution of c_n not available. Control limits h were obtained via simulation.

The MEWMC procedure

(1) Multistandardize X_n into U_n

(2) For each reading n, compute the matrix $\mathbf{S_n}$

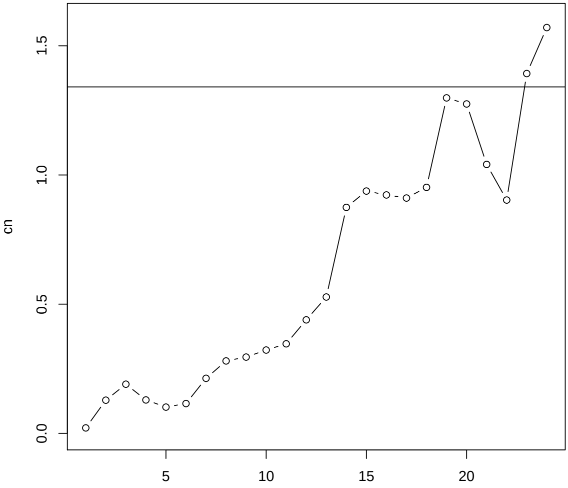
(3) Compute the statistic c_n

(4) For each successive readings n, plot this quantity c_n on a control chart, and the MEWMC chart gives an out-of-control signal as soon as $c_n > h$

Ambulatory Monitoring Data

- Concern is with physiological variables SBP, DBP, MAP, HR.
  - SBP: A mean systolic blood pressure
  - DBP: A mean diastolic blood pressure
  - MAP: An overall mean arterial pressure
  - HR: A mean of heart rate

- Use the MEWMC for dispersion control.
- The in-control distribution of the readings is assumed to be known exactly from the historical sequence. (Parameters known)
- Set λ to 0.1 and the IC ARL to 500.
- Control limit is $h = 1.3409$


cn
0.0
0.5
1.0
1.5
5
10
15
20

- MEWMC chart signals an OOC behavior at observation 22
- Shift apparently came around reading 5
- With these clues, we go back and diagnose
- The residual variances of MAP and HR have changed. The other variables seem fine.

Conclusion

- We define a new process, MEWMC
- We establish some theoretical properties
- We use the MEWMC to set up a method to monitor the covariance matrix of a process

Some References

[1] Chambers, J. M., (1971). Regression Updating. *Journal of the American Statistical Association*, **66**, 744–748

[2] Eaton, M. L.(1983) *Multivariate Statistics: A vector approach*. John Wiley and Sons New York.

[3] Hawkins, D. M., Maboudou-Tchao, E. M., 2008 Multivariate exponentially weighted moving covariance matrix. Technometrics, 50, 155-166

[4] Hawkins, D. M. and Eplett, W. J. R. (1982) "The Cholesky Factorization of the inverse correlation or covariance matrix in multiple regression". *Technometrics*. 24, 191-198.

[5] Montgomery, D. C. (2005), *Introduction to Statistical Quality Control*. 5th edition, Wiley