Departamento de Estadística, Facultad de Ciencias
Universidad Nacional de Colombia

Una carta de control multivariada CUSUM basada en antirangos

Camilo Uzaheta
bcuzahetab@unal.edu.co

Diciembre 13, 2018

Content

Entre las metodologías existentes para el monitoreo de cartas de control multivariadas libres de distribución se encuentran:

- Usar CM-EWMA, con ($\lambda < 0.05$). Pero se vuelve ineficiente para detectar cambios relativamente grandes.
- Realizar una transformación apropiada para llevarla a la normalidad multivariada.
- Cartas de control multivariadas no paramétricas basadas en ordenamiento.
  - Rango longitudinal
  - Rango transversal en las componentes
- Aprendizaje estadístico (k-chart, Sun & Tsung (2003))

Entre las cartas de control basadas en ordenamiento:

Rango longitudinal

Es el ordenamiento entre observaciones en diferentes puntos de tiempo.

- Carta multivariada no paramétrica del signo.
- Carta multivariada no paramétrica del rango signado.
- Carta multivariada no paramétrica basada en el signo espacial.
- Carta multivariada no paramétrica basada en el rango espacial.
- Carta multivariada no paramétrica basada en el concepto de profundidad.

Rango transversal en las componentes

El ordenamiento se puede dar entre diferentes componentes de un vector observado en un punto de tiempo

- Multivariate Antirank CUSUM (MA-CUSUM) Qiu & Hawkins (2001)
- Hwang Multivariate Antirank CUSUM (HMA-CUSUM) Hwang (2016)

Antirango

Por definición, el antirango de una muestra aleatoria $X_1, X_2, ..., X_p$ es una permutación A del vector de índices $(1, 2, ..., p)$ tal que:

$$X_{A_1} \leq X_{A_2} \leq \cdots \leq X_{A_p}$$

donde estas son las estadísticas de orden de la muestra.

A modo de ejemplo, sea el vector X

$$X = (-1, 5, 0, 3, 1, -2)$$

el vector de rangos asociados de X

$$R = (2, 6, 3, 5, 4, 1)$$

y el vector de antirangos de X

$$AR = (6, 1, 3, 5, 4, 2)$$

¿Por qué usar el antirango?

Estamos interesados en detectar un cambio en alguna de las componentes.
Para solucionar este problema, el uso del antirango es particular eficiente en detectar posible decrementos o incrementos, en alguna de las componentes sin saber cuál puede ser.

Sea $\boldsymbol{X}(n) = (X_1(n), X_2(n), ..., X_p(n))^T$ una observación p-dimensional de la característica de control durante la fase II. ($\boldsymbol{\mu}_0 = 0$)
Sea $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_p)^T$ el vector de medias en el tiempo n, la hipótesis nula de interés para el monitoreo de la media del proceso:

$$H_0 : \mu_1 = \mu_2 = ... = \mu_p = 0$$

Siendo equivalente a

$$H_0^{(1)} : \mu_1 = \mu_2 = ... = \mu_p$$

y

$$H_0^{(2)} : \sum_{j=1}^{p} \mu_j = 0$$

Sea $\boldsymbol{A}(n)$ el vector de antirangos del vector $\boldsymbol{X}_n$

$$\boldsymbol{A}(n) = (A_1(n), A_2(n), ..., A_p(n))^T$$

Por ahora, consideremos el primer antirango $A_1(n)$. Se va a definir el vector $\boldsymbol{\xi}_1(n) = (\xi_{1,1}(n), \xi_{1,2}(n), ..., \xi_{1,p}(n))$, en donde, para $1 \leq j \leq p$, se define

$$\xi_{1,j}(n) = \begin{cases} 1, & \text{si } A_1(n) = j \\ 0, & \text{e.o.c} \end{cases} \tag{1}$$

El vector $\boldsymbol{\xi}_1(n)$, es una indicadora, la cual indica el valor mas pequeño de las p componentes del vector observado en el tiempo n.

Bajo $H_0^{(1)}$, suponga que $E(\xi_{1,j}(n)) = g_{1,j}$ para $j = 1, 2, ..., p$. Luego, la distribución de probabilidad de $A_1(n)$ es $\boldsymbol{g}_1 = (g_{1,1}, g_{1,2}, ..., g_{1,p})^T$. Qiu & Hawkins (2001) muestran que bajo algunas condiciones de regularidad, cambios que violen $H_0^{(1)}$ resultan en un cambio de la distribución de $A_1(n)$. Entonces, evaluar $H_0^{(1)}$ es equivalente a

$$H_0^{(1)*} : \text{La distribución de } A_1(n) \text{ es } \boldsymbol{g}_1$$

Para evaluar $H_0^{(1)*}$, realizaron una combinación de un test para la bondad de ajuste para una función de distribución y la idea de Crosier (1988) para construir un procedimiento CUSUM multivariado.

Carta MA-CUSUM

Qiu & Hawkins (2001)

Sea $\boldsymbol{U}_0^{obs} = \boldsymbol{U}_0^{esp} = \boldsymbol{0}$ dos vectores de dimensión p, y $k_1 \geq 0$ es una constante.

$$\begin{cases} \boldsymbol{U}_n^{obs} = \boldsymbol{0}, & \text{si } B_n \leq k_1 \\ \boldsymbol{U}_n^{esp} = \boldsymbol{0}, & \\ \boldsymbol{U}_n^{obs} = (\boldsymbol{U}_{n-1}^{obs} + \xi_1(n))(1 - k_1/B_n), & \text{si } B_n > k_1 \\ \boldsymbol{U}_n^{esp} = (\boldsymbol{U}_{n-1}^{esp} + \boldsymbol{g}_1)(1 - k_1/B_n) & \end{cases}$$

donde

$$B_n = \left\{\left(\boldsymbol{U}_{n-1}^{obs} - \boldsymbol{U}_{n-1}^{esp}\right) + (\xi_1(n) - \boldsymbol{g}_1)\right\}^T \left(diag(\boldsymbol{U}_{n-1}^{esp} + \boldsymbol{g}_1)\right)^{-1} \left\{\left(\boldsymbol{U}_{n-1}^{obs} - \boldsymbol{U}_{n-1}^{esp}\right) + (\xi_1(n) - \boldsymbol{g}_1)\right\}$$

Se define

$$C_n = \left(\boldsymbol{U}_n^{obs} - \boldsymbol{U}_{n-1}^{esp}\right)^T \left(diag(\boldsymbol{U}_n^{esp})\right)^{-1} \left(\boldsymbol{U}_n^{obs} - \boldsymbol{U}_{n-1}^{esp}\right)$$

o de forma similar

$$C_n = max(0, B_n - k_1)$$

La carta de control genera una señal, indicando una violación de $H_0^{(1)*}$, si

$$C_n > h_1 \qquad (2)$$

Es posible construir la carta MA-CUSUM basada en mas componentes del vector de antirangos. Sea $(A_{j_1}(n), A_{j_2}(n), ..., A_{j_q}(n))^T$ cualquier vector de q componentes del vector de antirangos con $j_1 < j_2 < ... < j_q$. Asumiendo que la distribución de $(A_{j_1}(n), A_{j_2}(n), ..., A_{j_q}(n))^T$ cuando $H_0^{(1)}$ es verdadera esta determinada en el espacio muestral:

$$S(j_1, j_2, ..., j_q) = \{(j_1, j_2, ..., j_q) : i_1, i_2, ..., i_q \text{ son } q \text{ enteros diferentes en } (1, 2, ..., p)\}$$

El espacio muestral $S(j_1, j_2, ..., j_q)$ cuenta con $P_{q,p} = p(p-1)\dots(p-q+1)$ elementos.

- La distribución $\boldsymbol{g}_{.}$ bajo $H_0^{(1)}$, se puede calcular con datos previos con las frecuencias relativas. (Algebraicamente o computacionalmente si se conoce la distribución conjunta)
- Llevar a cabo simulaciones de este procedimiento, se puede hacer a partir de simular variables aleatorias de una multinomial. (Libre de distribución)
- La distribución $\boldsymbol{g}_{.}$ tiene en cuenta la asociación entre las variables.
- No detectar cambios cuando los componentes del vector de medias son todas iguales pero distintas a cero.
- No es eficiente para detectar cambios grandes en el vector de medias.

Carta HMA-CUSUM
Hwang (2016)

El procedimiento de la carta HMA-CUSUM es igual al procedimiento de la carta MA-CUSUM, el cambio se da en el vector indicadora.

Para $\xi_{1,2,\dots,q_1}(n)$, $q_1 \leq p-1$

$$\xi_{(1,2,\dots,q_1),j}(n) = \begin{cases} |X_{A_1(n)}(n)|, & \text{si } (A_1(n), A_2(n), ..., A_{q_1}(n))^T = \\ & \text{j-ésimo elemento de } S(1,2,...,q_1) \\ 0, & \text{e.o.c} \end{cases}$$

Para $\xi_{q_2,..,p-1,p}(n)$, $q_2 \geq 2$

$$\xi_{(q_2,..,p-1,p),j}(n) = \begin{cases} |X_{A_p(n)}(n)|, & \text{si } (A_{q_2}(n), ..., A_{p-1}(n), A_p(n))^T = \\ & \text{j-ésimo elemento de } S(q_2,..,p-1,p) \\ 0, & \text{e.o.c} \end{cases}$$

Esta aplicación es tomada de Hwang (2016). Se generaron 5000 observaciones en control de la distribución beta multivariada

$$y_{11}, y_{12}, y_{13}, y_{14}, y_{15} \sim \mathrm{BETA}\left(\mu = \frac{\exp(1 - x_{11} - 0.25x_{12})}{1 + \exp(1 - x_{11} - 0.25x_{12})}, \phi = 10\right).$$

Y los residuales deviance $d_{11}, d_{12}, d_{13}, d_{14}, d_{15}$ fueron calculados basados en la regresión beta.

Table VI. Average run lengths for the shift levels of y_{11}

Shift level	Original approach	Proposed approach	Hotelling's approach
0	165.12 (3.47)	165.88 (4.27)	163.81 (2.29)
−0.05	82.47 (1.79)	29.83 (0.83)	154.72 (2.22)
−0.1	24.58 (0.49)	6.57 (0.15)	75.48 (1.05)
−0.2	6.74 (0.08)	2.18 (0.03)	24.28 (0.33)

Bibliografia I

Crosier, R. B. (1988), 'Multivariate generalizations of cumulative sum quality-control schemes', *Technometrics* **30**(3), 291–303.

Hwang, W.-Y. (2016), 'A new rank-based multivariate cusum approach for monitoring the process mean', *Quality and Reliability Engineering International* **32**(3), 1167–1178.

Qiu, P. & Hawkins, D. (2001), 'A rank-based multivariate cusum procedure', *Technometrics* **43**(2), 120–132.

Sun, R. & Tsung, F. (2003), 'A kernel-distance-based multivariate control chart using support vector methods', *International Journal of Production Research* **41**(13), 2975–2989.

Gracias!