



Est. María. I. Flury
Est. Cristina A. Barbiero
Est. Marta Ruggieri

Instituto de Investigaciones Teóricas y Aplicadas. Escuela de Estadística.

PROPUESTAS PARA LA DETERMINACIÓN DE LOS PARÁMETROS DEL GRÁFICO DE CONTROL MEWMA

INTRODUCCIÓN

En la actualidad es común el uso de técnicas de control estadístico del proceso (SPC) para el monitoreo de una única característica de calidad. Los gráficos utilizados con este fin son los SHEWHART, CUSUM, EWMA, etc..

Con la aparición de modernos equipos tomadores de datos y de computadoras conectadas en línea, resulta factible monitorear simultáneamente varias características de calidad correlacionadas. Este control se hace posible mediante el SPC multivariante usando gráficos como el T^2 de HOTELLING, el MCUSUM y el MEWMA.

El más utilizado y desarrollado hasta el momento es el T^2 . Sin embargo, existen otras propuestas que incorporando la información pasada del proceso, permiten aumentar la potencia del gráfico ante salidas de control.

GRÁFICO MULTIVARIADO DE PROMEDIOS MÓVILES PONDERADOS EXPONENCIALMENTE (MEWMA)

Este gráfico, a diferencia del T^2 de Hotelling, es un gráfico ponderado en el tiempo, en el cual cada punto graficado contiene información no sólo del último período sino también de los anteriores. A cada período se le da un peso que decrece en forma exponencial a medida que se aleja del actual.

Fue desarrollado por C.Lowry y colaboradores en el año 1992, quienes lo definieron como una extensión del EWMA univariado.

El interés se centra en el control simultáneo de p características de calidad correlacionadas entre sí. En este caso $\mathbf{X}_1, \mathbf{X}_2, \dots$ son vectores de longitud p . Estos vectores pueden representar observaciones individuales o medias de subgrupos tomados del proceso.

Para estudiar la performance de este gráfico se supone que los vectores aleatorios \mathbf{X}_i son independientes y están idénticamente distribuidos según una normal p -variante con vector de medias $\boldsymbol{\mu}_i$ y matriz de covariancias $\boldsymbol{\Sigma}_x$.

Por simplicidad se supone que $\boldsymbol{\Sigma}_x$ es conocida. En la práctica es necesario recolectar datos durante un cierto tiempo en el que el proceso está bajo control, para estimar $\boldsymbol{\Sigma}_x$. Estos datos también pueden usarse para chequear los supuestos de normalidad e

independencia. Si éstos no se cumplen, las propiedades del gráfico pueden verse afectadas y las señales de fuera de control podrían carecer de significado..

Sin pérdida de generalidad se supone que el vector de medias cuando el proceso está



bajo control es $\mu_0 = (0,0,0,\dots,0)'$.

Si $\mu = \mu_0$ el proceso está bajo control.

Si $\mu \neq \mu_0$ el proceso está fuera de control.

Tomando como vector de partida a $Z_0 = \mathbf{0}$, los vectores MEWMA se definen como :

$$Z_i = \mathbf{R} X_i + (\mathbf{I} - \mathbf{R}) Z_{i-1} \quad \text{para } i \in \mathbf{N}; \mathbf{R} = \text{diag} (r_1, r_2, \dots, r_p);$$

$$0 < r_j \leq 1 \quad ; j = 1, 2, \dots, p$$

Si todos los r_j son iguales, los vectores MEWMA pueden reescribirse de la siguiente forma:

$$Z_i = rX_i + (1 - r) Z_{i-1}$$

Los r_j marcan la profundidad de la memoria para cada variable. Cuanto mayor sea r_j menor será la profundidad, es decir menor peso tendrán las observaciones anteriores.

Se pueden usar valores de r_j específicos para cada variable. Aquí es donde se diferencia un análisis direccional de uno sin dirección específica. Si se usa el mismo peso r_j para todas las variables, el gráfico es de dirección invariante, debido a que una señal fuera de control no podrá ser atribuida a alguna variable en especial ya que todas tienen igual peso. Si se da distinto peso a las variables, el gráfico será de dirección específica y en ese caso, una señal de fuera de control será atribuida a los valores específicos de aquella variable a la que se le otorgó el mayor peso. Cuando se utiliza un r común y éste es igual a 1, se obtiene el gráfico T^2 de Hotelling.

En cualquier caso el estadístico "MEWMA" que se grafica es:

$$T_i^2 = Z_i \sum_z^{-1} Z_i$$

donde \sum_z^{-1} es la inversa de la matriz de covariancias de Z_i que se obtiene como:

$$\sum_z = \frac{r}{2-r} [1 - (1-r)^{2i}] \sum_x$$

La señal de salida de control se produce cuando T_i^2 supera un cierto valor h ($h > 0$), siendo h el límite de control elegido para conseguir un valor determinado de ARL cuando el proceso está bajo control (ARL_0).

Con las siglas ARL se representa a la longitud de corrida promedio, la cual es una característica usada con frecuencia para evaluar el comportamiento de un diagrama de control. Es el número de puntos que en promedio deberán graficarse entre salidas de control identificadas por el gráfico.

Cuando los r_j son iguales a r , la performance del ARL del MEWMA depende solamente del parámetro de no centralidad λ y no de la dirección del cambio que se desea detectar; siendo



$$\lambda = n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) .$$

Previamente a la utilización de un gráfico MEWMA, es necesario definir los parámetros **h** y **r** mencionados anteriormente, que lo caracterizan. Diferentes autores han propuesto métodos para su obtención.

1) Propuesta de C. Lowry y colaboradores

Lowry y col. (1992) obtuvieron mediante simulación diversas tablas que permiten hallar valores adecuados de **r** y de **h**. Ellos están en función del número *p* de variables consideradas, del ARL bajo control (ARL_0) y del cambio **d** que se desea detectar en el vector de medias. Además, las tablas proporcionan el ARL correspondiente al cambio **d** que se desea detectar.

Estas tablas no son completas pues están dadas para determinados valores de *p* (2, 3, 4, 5, 10 y 20) y para determinados valores de ARL_0 (50; 100, 200, 500, 1000) .

Para el ejemplo publicado en *Technometrics* por estos autores, donde se consideran datos de 10 subgrupos de tamaño $n = 1$, siendo $p = 2$ el número de variables analizadas, se obtendrían $h = 8,79$ y $r = 0,10$ cuando se desea un $ARL_0 = 200$. El valor de $ARL_d = 7,76$.

2) Propuesta de García-Díaz y Aparisi

Otra sugerencia para la obtención de los parámetros óptimos del gráfico MEWMA, es la de García-Díaz y Aparisi (2.003), quienes desarrollaron un programa que puede ser utilizado en cualquier situación.

El software propuesto fue programado en entorno Windows y está basado en el concepto de Algoritmos Genéticos. Estos son algoritmos de optimización en los cuales la búsqueda del óptimo global se realiza al pasar de una población inicial de individuos (generación), a otra población nueva (generación siguiente) mediante la aplicación de operadores genéticos.

La función $Z(x)$ a optimizar es del tipo:

$$Z(x) = \sum_{k=1}^K w_k z_k (x)$$

donde *K* es el número de objetivos a satisfacer y $w_k \geq 0$ es el peso o ponderación correspondiente al objetivo $z_k (x)$, que puede interpretarse como la importancia que el mismo tiene, en comparación con los demás. Este procedimiento convierte un problema de objetivos múltiples, en un problema de optimización con un único objetivo.

Al determinar la función, los autores se basaron en la propuesta de Woodall quien define tres regiones de interés para el diseño estadístico de los gráficos de control: región bajo control, región de indiferencia y región de fuera de control. Las mismas toman como base la magnitud del cambio **d** que desea detectarse en el vector de medias y están delimitadas por dos valores *A* y *B*, de acuerdo a lo siguiente:

. **Región "bajo control"** $[0, A]$: en esta región, el cambio en el vector de medias del



proceso varía de $d = 0$ a $d = A$. Este cambio no se considera de interés y por tanto no se lo desea detectar. En esta zona se requiere un ARL máximo, ya que si el gráfico mostrara una señal de fuera de control, correspondería a una falsa alarma.

. **Región "de indiferencia" (A, B):** cubre los valores de $A < d < B$. En esta región es indiferente detectar o no, el cambio en el proceso.

. **Región "fuera de control" [B, ∞):** cubre los valores de $d > B$. En esta área se requiere la máxima potencia para detectar el cambio, necesitando un ARL mínimo.

Relacionando estas regiones con la función a optimizar, se tiene que los dos objetivos a considerarr son:

- 1) $z_1(x)$: se refiere a la diferencia entre el ARL del proceso bajo control que corresponde al gráfico seleccionado (ARL_0) y el ARL mínimo que se desea cuando $d=0$ (ARL_{\min}). Este es un objetivo a maximizar, ya que se desea obtener gráficos de control que satisfagan la relación $ARL_0 \geq ARL_{\min}$.
- 2) $z_2(x)$: se refiere al valor del ARL para un punto $d = B$ (ARL_B). Este es un objetivo a minimizar, ya que ARL_B debe ser mínimo.

Con estos dos objetivos, el problema de optimización consiste en:

$$\begin{aligned} \text{Maximizar } Z(x) &= \sum_{k=1}^K w_k z_k(x) = w_1 z_1(x) + w_2 z_2(x) \\ &= w_1 (ARL_0 - ARL_{\min}) - w_2 ARL_B \end{aligned}$$

Los valores de w_1 y w_2 son subjetivos. En su publicación, García-Díaz y Aparisi dan a w_1 valores comprendidos entre 1 y 10 y a w_2 valores comprendidos entre 5 y 400. Si bien el programa admite cualquier relación entre w_1 y w_2 , los autores mencionados tomaron siempre en cada aplicación, valores de w_1 inferiores a los de w_2 .

Para la utilización del programa propuesto, es necesario dar los siguientes datos:

- . ARL deseado cuando el proceso se encuentra bajo control
- . número de variables considerado
- . tamaño de muestra o subgrupo del gráfico de control
- . distancia **d**
- . datos realacionados con el Algoritmo Genético, entre ellos los valores de w_1 y w_2 .

El programa proporciona valores por default, los que pueden modificarse.

A partir de estos datos, el programa proporciona los valores del límite superior de control **h**, la constante de suavizamiento o profundidad de la memoria **r** óptima y los valores de ARL obtenidos cuando el proceso está bajo control y cuando el mismo se aparta en una magnitud **d**.

García-Díaz y Aparisi compararon los valores que se obtienen con el programa propuesto, con los determinados por Lowry y col.(1992) y por Prabhu y Ranger (1997). Hallaron que los mismos son muy similares y que, al igual que lo habían indicado esos autores,



el parámetro r aumenta, a medida que lo hace el tamaño del cambio que desea detectarse.

Para el ejemplo mencionado en el punto anterior, se obtendría $h = 9,18$ y $r = 0,14$ cuando se utilizan ponderaciones $w_1 = 10$ y $w_2 = 50$. El valor de $ARL_d = 10,08$.

La ventaja de esta propuesta, es que la misma se adapta a cualquier valor de ARL deseado y cualquier número de variables involucrado, lo que la hace más general al momento de su aplicación.

3) Propuesta de Lau Meng Cheng y colaboradores

Los autores sugieren estimar el límite superior de control en los gráficos MEWMA a través de : $\bar{M} + 3S_M$, como el usado en la distribución normal tradicional , siendo:

$$\bar{M} : \text{promedio de los } M_i \quad \text{con} \quad M_i = T_i^2 = Z_i^T \sum_{zi}^{-1} Z_i;$$

S_M : desvío estándar de los M_i

Proponen un diagrama de flujo para el cálculo y trazado de la estadística MEWMA según se consideren observaciones individuales o datos agrupados.

Los algoritmos comienzan con un conjunto de datos multivariados con subgrupos en un número: $m \geq 10$: El tamaño muestral es n y decide el tipo de gráfico, ya sea con datos agrupados (A) u observaciones individuales (B).

El siguiente diagrama de flujo (adaptado del artículo original) indica la posible selección del gráfico de control multivariado:

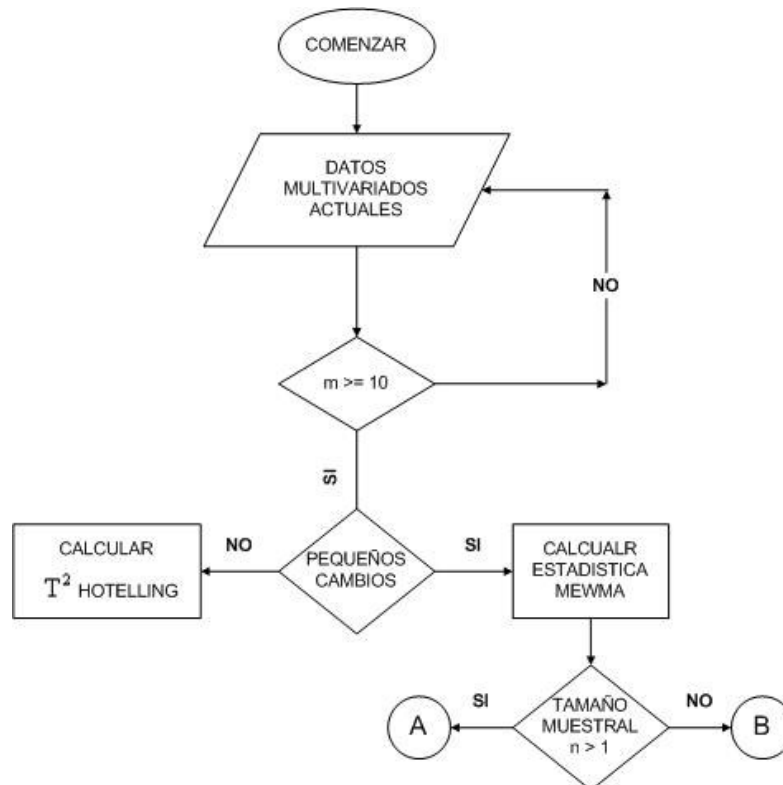




Diagrama para $n > 1$

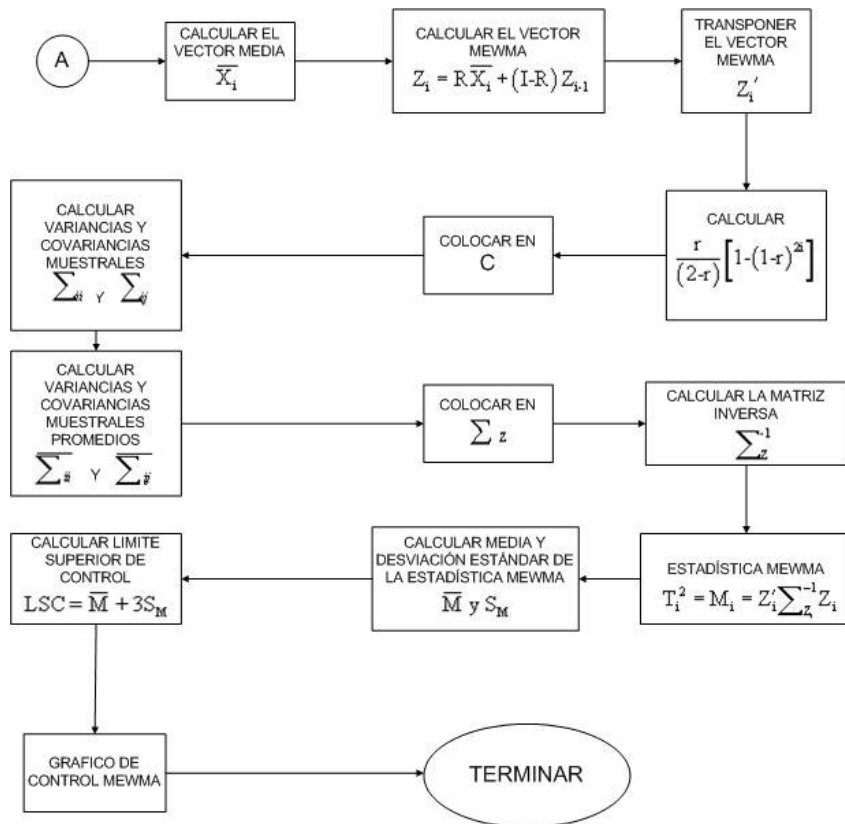
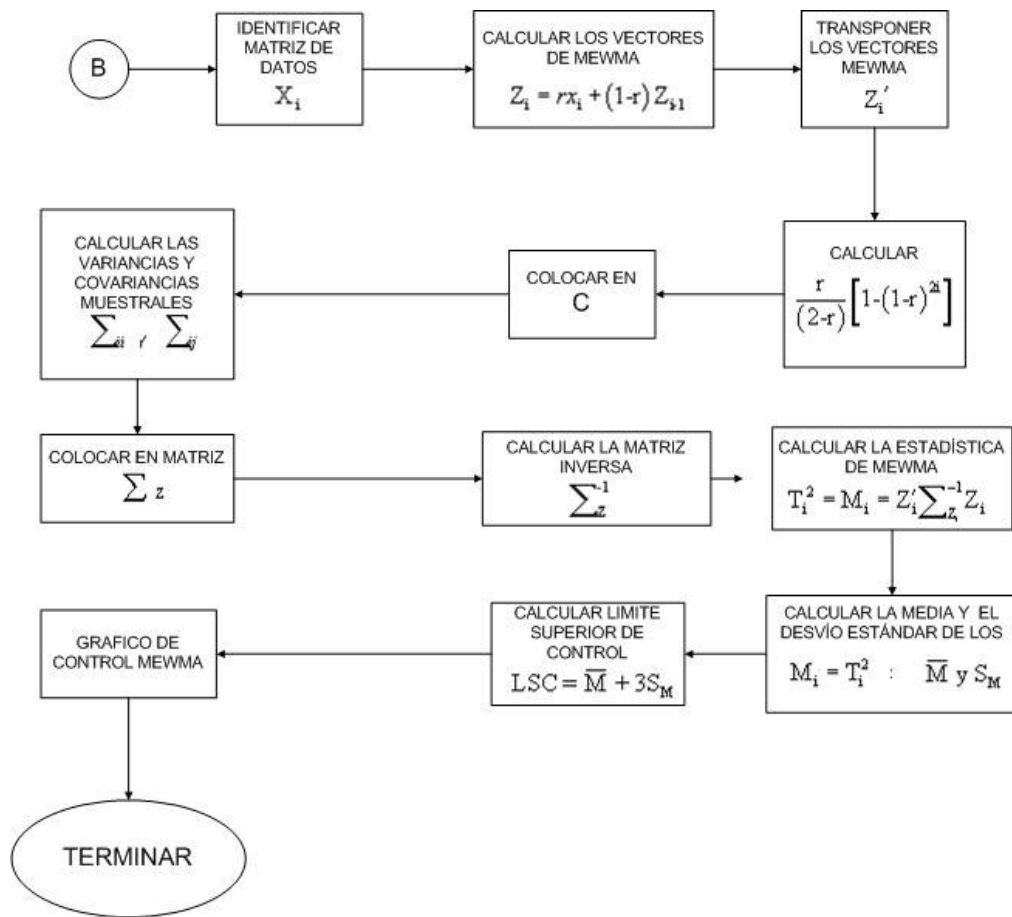




Diagrama de flujo para n=1





Obsérvese que para ambos algoritmos **A** y **B** se necesita el cálculo de la inversa de la matriz de variancias y covariancias de los datos originales y además el de \sum_{zi}^{-1} (matriz inversa de variancias y covariancias de los vectores MEWMA Z_i).

El (k, l) -ésimo elemento de $\sum_{zi}^{(k,l)} = \frac{r}{2-r} [1 - (1-r)^{2i}] \sum_x$; si $r_1 = r_2 = \dots = r_p = r$.

Los algoritmos propuestos en **B** se verificaron con los datos presentados por Lowry y col., para facilitar comparación con las opciones dadas anteriormente. Suponiendo un valor de $r = 0,10$ (dato), se obtiene un $h = 16,36$.

Si bien los autores no lo mencionan explícitamente en su artículo, es de suponer que este límite se utilizaría como valor de referencia para futuros procesos.

El valor de r no se determina a través de este procedimiento sino que tiene que haber sido previamente fijado. Tampoco se hace mención al valor de ARL_0 ni al de la magnitud del cambio d de que se desea detectar.

CONSIDERACIONES FINALES

En la literatura científica se han presentado distintas alternativas, pero no todas ellas conducen a los mismos resultados.

El programa desarrollado por Gómez-Díaz y Aparisi proporciona resultados similares a los indicados en las tablas de Lowry y colaboradores. En cambio la metodología propuesta por Lau Meng Cheng y colaboradores, basada en datos del proceso, conduce a valores diferentes para límites de control.

Por lo expuesto, y dada la importancia que tiene la definición del límite de control a la hora de implementar una estrategia MEWMA, se evidencia la necesidad de profundizar el tratamiento de este tema.

REFERENCIAS BIBLIOGRÁFICAS

- Aparisi, F y García-Díaz, J.C. (2004): "A Multiobjective Optimization for the Ewma and Mewma Quality Control Charts". Inverse Problems. Design and Optimization Symposium. Rio de Janeiro, Brazil, 2004.
- García-Díaz, J.C. y Aparisi, F. (2003): "Optimización de los Gráficos de Control Estadístico de Procesos Ewma y Mewma mediante Algoritmos Genéticos". 27 Congreso Nacional de Estadística e Investigación Operativa. Lleida, 8-11 de abril de 2003.
- Lau Meng Cheng, Yuwaldi Away, Mohammad Khatim Hasa. Department of Industrial Computing. Faculty of Information Science and Technology. National University of Malaysia.
- Lowry, C.A.; Woodall, W.H.; Champ, C.W. and Riigdon, S.E. (1992): "A Multivariate Exponentially Weighted Moving Average Control Chart". Technometrics, 34(1), 46-53.
- Prabhu, S.S. and Runger, G.C. (1997): "Designing a Multivariate EWMA Control Chart". Journal of Quality Technology, 29(1), 8-15.

