



Flury, María Isabel

Quaglino, Marta

Barbiero, Cristina

Ruggieri, Marta

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística

PROPUESTA DE IMPLEMENTACIÓN DE UN GRÁFICO MULTIVARIADO DE PROMEDIOS MÓVILES PONDERADOS EXPONENCIALMENTE (MEWMA) PARA VARIABLES CON COMPORTAMIENTO ASIMÉTRICO.

1. Introducción

A través del tiempo se ha ido intensificando el interés por el desarrollo y aplicación de las técnicas de Control Multivariado de Procesos (MSPC). Ellas toman en consideración no solamente el valor individual de cada una de las variables que son controladas en un proceso, sino también la relación existente entre las mismas, permitiendo generar mecanismos de control más sensibles a causas asignables de variación.

Los diagramas o gráficos de control son una de las herramientas más utilizadas para detectar rápidamente cualquier alteración que pueda ocasionar un cambio indeseado en la calidad del producto. Fueron diseñados originalmente por W. Shewhart para el caso univariado. Posteriormente se desarrollaron gráficos multivariados: el basado en la estadística T^2 de Hotelling (1947), a veces denominado gráfico Shewhart multivariado; el Gráfico Multivariado de Sumas Acumuladas (MCUSUM) con propuestas de Woodall y Ncube en 1985 y de Crosier en 1986 y el Gráfico Multivariado de Promedios Móviles Ponderados Exponencialmente (MEWMA), propuesto por Lowry y colaboradores en 1992. El gráfico T^2 está basado en la observación más reciente, siendo insensible a cambios pequeños a moderados en el vector de medias del proceso. Comparativamente, el diseño de la estrategia T^2 es mucho más sencillo, ya que la definición del límite de control se establece a partir de un percentil de una distribución de probabilidad conocida. En cambio, tanto en el MEWMA como en el CUSUM MULTIVARIADO, el límite se establece a partir de procedimientos más complejos (simulación o análisis por cadenas de Markov) que tienen en cuenta el cumplimiento de ciertas propiedades que el usuario debe establecer. Para ambos gráficos existen tablas para determinar el límite de control, las que no siempre abarcan todos los casos. También se han desarrollado programas más generales, los cuales no son fácilmente accesibles

La mayoría de los trabajos publicados sobre estos tres diagramas de control se basa en el supuesto de que, cuando las variables de calidad son continuas, la distribución subyacente de las observaciones es Normal Multivariada, porque la misma se adapta a muchas situaciones prácticas. Por ello, para garantizar una adecuada elección de la estrategia, es necesario chequear previamente la validez del supuesto de Normalidad.

La no Normalidad no es un problema muy serio cuando se trabaja con subgrupos de tamaño grande, ya que el Teorema Central del Límite asegura que el vector de medias es aproximadamente multinormal para distintas distribuciones multivariadas. Sin embargo, en la industria es frecuente el uso de subgrupos de tamaño pequeño, incluso de tamaño uno, por



lo cual se hace necesario chequear el supuesto y analizar la sensibilidad del gráfico ante alejamientos de la normalidad, previamente a su implementación.

Cuando las variables continuas responden a otros modelos porque tienen por ejemplo un comportamiento asimétrico, deberá elegirse otra distribución que las represente. La distribución Gamma, que es asimétrica hacia la derecha, puede ser una alternativa adecuada. En esta distribución, que está definida para variables que toman valores positivos, los valores pequeños se presentan con gran frecuencia mientras que los grandes se presentan ocasionalmente. En el ámbito industrial, el modelo Gamma se ha utilizado en problemas de confiabilidad y mantenibilidad para estudiar por ejemplo el tiempo que requiere el mantenimiento de una máquina o el tiempo hasta el fallo de un sistema o componente. También ha sido de utilidad para representar la resistencia de ciertos componentes o el tiempo total que se requiere para completar una operación, cuando ésta se lleva a cabo en distintas subestaciones. Su uso también es frecuente en problemas de tiempos de espera, de tráfico en líneas telefónicas o cuando se analizan datos referidos a ingresos familiares o a precipitaciones.

El problema del comportamiento del MEWMA bajo distribución Gamma multivariada fue abordada por diversos autores, entre ellos Stoumbos y Sullivan (2002). Continuando esta línea de investigación, en su trabajo de Tesis, Flury (2010) efectúa un análisis alternativo y compara la eficiencia de este diagrama bajo ambas propuestas cuando el proceso está bajo control y cuando se aparta de la situación de control. En ambos trabajos se supone la misma estructura para la matriz de variancias y covariancias de las variables involucradas y se simulan de idéntica manera las situaciones fuera de control, a fin de facilitar comparaciones.

A partir de un Convenio de Cooperación firmado con una empresa metalúrgica del Gran Rosario, entre otras tareas, fue posible obtener información sobre distintas variables involucradas en diferentes procesos. Esta empresa fabricaba artículos en lotes, a través de múltiples subprocesos encadenados: corte de una barra de acero, doblado, tratamiento del trozo de materia prima con distintos tornos, cepillado, forjado, laminado, etc. Cada pieza requería de distinto número y tipo de operaciones y para fabricar cualquiera de ellas, las tareas combinaban el uso automático de maquinarias y la intervención de operarios especializados, probablemente con distinto grado de capacitación, según el procedimiento que debía aplicarse.

Como tema de Investigación, dentro del Proyecto "Métodos estadísticos para la mejora de la calidad y productividad", se pensó en evaluar la posibilidad de diseñar una adecuada estrategia MSPC para el control de los tiempos de producción de lotes de diferentes artículos, tomando datos reales del proceso. En particular se analizó la alternativa de implementación de un gráfico MEWMA para el control de los tiempos de producción de lotes de una pieza en la cual estaban involucrados tres procesos consecutivos. Si bien los subprocesos podrían considerarse independientes, la idea de plantear una estrategia multivariada se basó en conservar las tres mediciones simultáneamente como un indicador de eficiencia de producción de los lotes. Ello posibilitaba mantener la información sobre cada subproceso, la cual se hubiera perdido si el control se planteaba por ejemplo sobre "el tiempo total de producción del lote". Las variables analizadas presentaban un comportamiento marginal asimétrico, por lo que la distribución Gamma podía considerarse como alternativa.



En el presente trabajo se analizan las dificultades encontradas en la implementación del gráfico mencionado, se sugieren valores para los dos parámetros que requiere su implementación (límite de control y constante de suavizamiento), se compara esta estrategia con la del gráfico T^2 convencional y se analizan diferencias con otros estudios efectuados.

2. Metodología: Gráfico MEWMA

La estrategia *MEWMA* (C. Lowry et al, 1992) es una extensión del *EWMA* univariado. Los vectores de p características de calidad pueden representar observaciones individuales o medias de subgrupos de tamaño n , tomados a intervalos regulares durante el proceso; y el conjunto de parámetros μ_0 y Σ_0 pueden ser conocidos o estimados a partir de un período normal del proceso.

Trabajando con vectores centrados en su media (el vector de medias cuando el proceso está bajo control es el vector nulo), los vectores *MEWMA* se definen como:

$$\vec{Z}_i = \mathbf{R}_{p \times p} \vec{X}_i + (\mathbf{I}_{p \times p} - \mathbf{R}_{p \times p}) \vec{Z}_{i-1},$$

para $i \in N$; $\mathbf{R}_{p \times p} = \text{diag}(r_1, r_2, \dots, r_p)$; $0 < r_j \leq 1$; $j = 1, 2, \dots, p$, con $\vec{Z}_0 = \vec{0}$.

Las constantes r_j representan la profundidad de la memoria y son conocidas como "constantes de suavizamiento" para cada variable. Si todas las r_j son iguales, los vectores *MEWMA* pueden reescribirse como:

$$\vec{Z}_i = r \vec{X}_i + (1-r) \vec{Z}_{i-1}.$$

Cuanto mayor sea r_j menor será la profundidad de la memoria del gráfico, es decir menor peso tendrán las observaciones anteriores, en cada medición Z_i . Si se usa el mismo peso r_j para todas las variables, el gráfico es de dirección invariante y una señal de salida de control no podrá ser atribuida a alguna variable en especial.

El estadístico *MEWMA* que se grafica, es la forma cuadrática:

$$T^2_i = Z_i \Sigma_z^{-1} Z_i,$$

donde la expresión asintótica de Σ_z , es $\Sigma_z = \frac{r}{2-r} \Sigma_0$.

La señal de salida de control se produce cuando la estadística T^2_i supera un cierto valor h positivo, que se define de modo de conseguir un valor determinado de *ARL* (*Longitud de corrida promedio*) cuando el proceso está bajo control. Cuando los r_j son iguales, la performance del *ARL* del *MEWMA* depende del parámetro λ de no centralidad de la estadística T^2 que se expresa como $\lambda = n(\mu_1 - \mu_0) \Sigma_0^{-1} (\mu_1 - \mu_0)$, donde μ_1 es el vector de medias que difiere de μ_0 en al menos una de las medias.



Stoumbos, Z. y Sullivan, J., (2002) evaluaron el desempeño del MEWMA en el caso de observaciones provenientes de distribuciones multivariadas Gamma, utilizando como medida de performance el ARL, simulando observaciones bajo control (con vector de medias nulo) y observaciones que responden a distintas situaciones fuera de control (con vector de medias no nulo). En el primer caso, el vector centrado se obtiene restando el vector de medias al vector de observaciones. Cuando se generan observaciones fuera de control, debe sumarse al vector centrado, el vector de cambio previsto. Específicamente, si \vec{Y}_i es un vector de observaciones con una distribución Gamma multivariada determinada y se utiliza una única constante de suavizamiento r , el vector MEWMA \vec{Z}_i para la i -ésima observación se calcula como:

$$\vec{Z}_i = r(\vec{Y}_i - \vec{\mu}_0 + \vec{\delta}_i) + (1-r)\vec{Z}_{i-1},$$

donde \vec{Y}_i es el vector de la i -ésima observación Gamma multivariada, $\vec{\mu}_0$ es el vector de medias cuando el proceso está bajo control y $\vec{\delta}_i$ es el vector con los cambios en las medias de las variables. Se supone $\vec{Z}_0 = \vec{0}$.

En el caso de la distribución Gamma, el vector de medias no puede especificarse independientemente de la matriz de covariancias. Si se utilizan observaciones Normales p -variadas para generar observaciones Gamma, y aquéllas tienen como vector de medias el vector nulo y como matriz de covariancias una matriz arbitraria Σ_X , los vectores gamma \vec{Y}_i generados, siguen una distribución Gamma p variada, con parámetro de forma $\alpha = n/2$, vector de medias $\vec{\mu}_Y = 0.5n d(\Sigma_X)$ y matriz de covariancias $\Sigma_Y = 0.5n \Sigma_X^{(2)}$, donde $d(\Sigma_X)$ es el vector cuyos elementos son los de la diagonal principal de Σ_X y $\Sigma_X^{(2)}$ es la matriz cuyos elementos son los de Σ_X elevados al cuadrado. Esto hace que todos los elementos de la matriz Σ_Y sean no negativos, con lo cual el método utilizado para generar observaciones Gamma multivariadas, no puede generar observaciones con correlaciones negativas. En el caso en el cual $\Sigma_X = \mathbf{I}$, resulta $\vec{\mu}_Y = 0.5n\mathbf{1}$ y $\Sigma_Y = 0.5n\mathbf{1}$, siendo $\mathbf{1}$ el vector de unos.

Esta situación es la que se consideraron Stoumbos y Sullivan en su trabajo y posteriormente M. Flury en su Tesis, a fin de comparar resultados. Stoumbos y Sullivan sugieren emplear como límite de control h para el gráfico de control, el mismo que se utiliza para distribuciones Normales. Recomiendan usar valores del parámetro de suavizamiento (r) menores a 0,05, ya que con ellos se obtiene un ARL bajo control, similar al del caso Normal. En el trabajo de Tesis, se determinan especialmente los límites de control mediante simulación, a través de la generación de variables gamma multivariadas. Los límites hallados son diferentes a los sugeridos por los autores mencionados y también los valores recomendados de la constante de suavizamiento r , especialmente cuando el número de variables involucradas es mayor que dos.

En esta propuesta de implementación del MEWMA no pudo utilizarse el mismo algoritmo para generar situaciones fuera de control, porque la relación entre la media y el desvío estándar de las distribuciones marginales no se mantiene constante.



3. Resultados

La información analizada en este trabajo se refiere a los tiempos de ejecución (por hora) de tres subprocesos de la pieza "bulón rosca cuadrada". La información acerca de cada proceso se recoge en "hojas de ruta" que acompañan a la producción de cada lote. En ellas se indica fecha del pedido, cantidad de piezas encargadas, fecha de realización de cada subproceso, nombre del operario y tiempo de ejecución. Los datos de las hojas de ruta son volcados a un sistema administrativo informatizado. Para el planteo del control estadístico fue necesario transformar los datos referidos a "tiempo total de ejecución de cada subproceso en un lote" a una medida comparable, eligiendo como tal el "tiempo unitario (en horas) por pieza, para cada subproceso".

Se obtuvo información para un total de 15 lotes, en un período de operación normal. Se ajustaron distribuciones Gamma a los datos de tiempos de cada subproceso y se estimaron los parámetros de forma y escala correspondientes (Tabla 1). Las estimaciones de medias, variancias y covariancias se muestran en la Tabla 2.

Tabla 1: Parámetros estimados y bondad de ajuste para cada subproceso

Estimaciones	Subproceso A	Subproceso B	Subproceso C
Parámetro de forma	1.906	3.581	2.314
Parámetro de escala	0.01045	0.0047	0.0035
p-value (bondad de ajuste)	> 0.25	> 0.25	> 0.25

Tabla 2: Estimaciones de Medias, Variancias y Covariancias de tiempos de subprocesos

Parámetro estimado	Subproceso A	Subproceso B	Subproceso C
Media	0.01861 hs.	0.00806 hs.	0.01689 hs.
Variancia	0.00019378	0.00007255	0.00002814
Covariancias	(A,B)=-0.00003305	(A,C)=0.00000575	(B,C)=0.00000621

Para elegir la constante r y el límite de control h , se estimaron curvas de ARL simulando observaciones de distribuciones Gamma con parámetros iguales a las estimaciones obtenidas. Como ya se mencionó, el algoritmo propuesto por Stoumbos y Sullivan (2002) no pudo utilizarse en este trabajo porque la relación entre media y desvío estándar de las distribuciones marginales no se mantiene constante

El límite de control h en cada caso fue fijado para obtener un ARL bajo control de 200. La Tabla 3 muestra los ARL estimados a partir de 5000 repeticiones de la variable aleatoria RL , para la situación bajo control y para cambios del vector de medias expresados en múltiplos de desviaciones estándar de cada variable. Específicamente para $\mu_1 = \mu_0 + \Delta$, con



$\Delta = (d^* \sigma_{11}, d^* \sigma_{22}, d^* \sigma_{33})$, para $d = 0.2$ a 2.4 , en incrementos de 0.2 . En la misma tabla se muestran los límites h utilizados.

El gráfico T^2 es un gráfico MEWMA con $r=1$. A fin de efectuar comparaciones, también se obtuvo para la estrategia T^2 el límite mediante el cual se obtiene un ARL bajo control de 200 y luego se determinó mediante simulación la curva ARL correspondiente. Se observó que los límites obtenidos difieren sustancialmente de los recomendados para el caso Normal (son mucho más altos), considerando igual p y el mismo ARL bajo control.

Analizando la tabla 3, se observa que a similar performance bajo control, los r mayores detectan más rápidamente las salidas de control uniformemente para cualquier desviación de la media (los ARL van disminuyendo). A partir de $r = 0.25$ la potencia del gráfico se estabiliza. El gráfico T^2 fue menos eficiente que cualquiera de las alternativas MEWMA consideradas para detectar los cambios, aún fijando especialmente el límite de control, considerando variables con distribución Gamma.

Tabla 3: ARL estimados para distintos r con estrategias MEWMA y T^2

d	Variantes de r en Gráficos MEWMA										T^2
	0,01	0,02	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,3	
0	200,00	199,96	199,11	204,20	201,90	209,18	199,93	199,35	203,25	202,60	202,5
0,2	198,00	198,00	196,00	200,00	198,00	207,44	190,19	179,62	164,16	158,68	168,1
0,4	197,13	190,43	184,90	177,46	168,18	149,82	130,86	119,83	118,46	117,08	131,5
0,6	172,64	144,46	124,76	114,32	106,04	86,09	79,67	73,02	73,55	72,16	101,6
0,8	146,22	109,55	89,47	75,13	68,22	51,39	44,95	44,20	42,49	43,83	76,6
1	123,38	86,12	66,45	55,24	48,19	33,09	28,84	26,39	26,41	27,46	59,8
1,2	104,68	70,09	52,70	42,78	36,50	23,76	19,62	17,34	17,58	17,02	42,8
1,4	90,08	58,81	43,25	34,86	29,65	18,41	14,45	12,78	12,04	11,80	31,6
1,6	78,58	50,05	36,55	29,26	24,58	14,86	11,57	9,91	9,06	8,55	33,5
1,8	69,16	43,57	31,68	25,08	21,05	12,41	9,51	7,93	7,18	6,70	16,9
2	61,73	38,49	27,73	21,91	18,28	10,78	8,06	6,74	5,98	5,50	12,1
2,2	55,54	34,30	24,69	19,57	16,24	9,43	7,06	5,81	5,12	4,73	8,7
2,4	50,49	31,05	22,23	17,54	14,60	8,42	6,26	5,15	4,49	4,07	6,2
h	575	388	286	228,5	192	113	84	68,95	59,71	53,5	31,4

Como análisis complementario, se muestran en la Figura 1 las curvas ARL obtenidas para valores de r pequeños, menores que $0,10$ y en la Figura 2, las correspondientes a valores de r más grandes. Se subdividen los valores de r en dos grupos para distinguir las variantes entre las curvas.



Figura 1: Curvas de ARL para estrategias MEWMA con $r < 0.10$

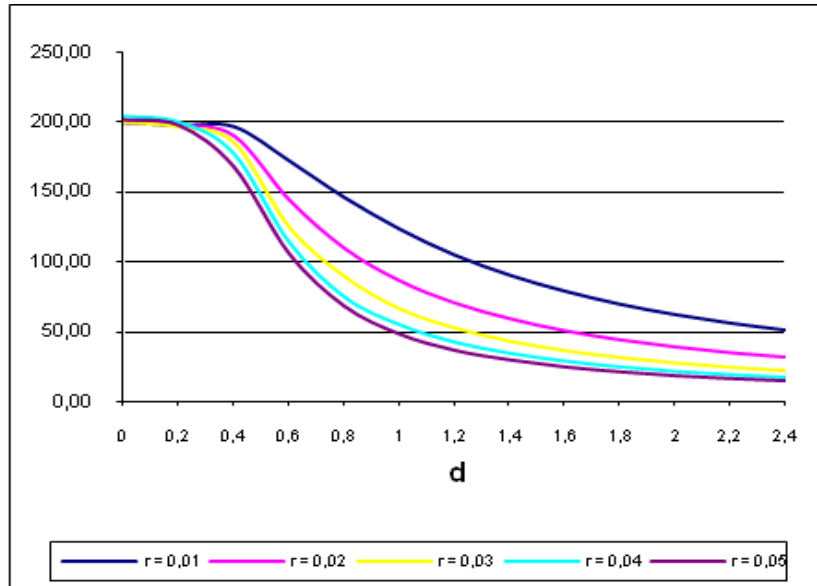
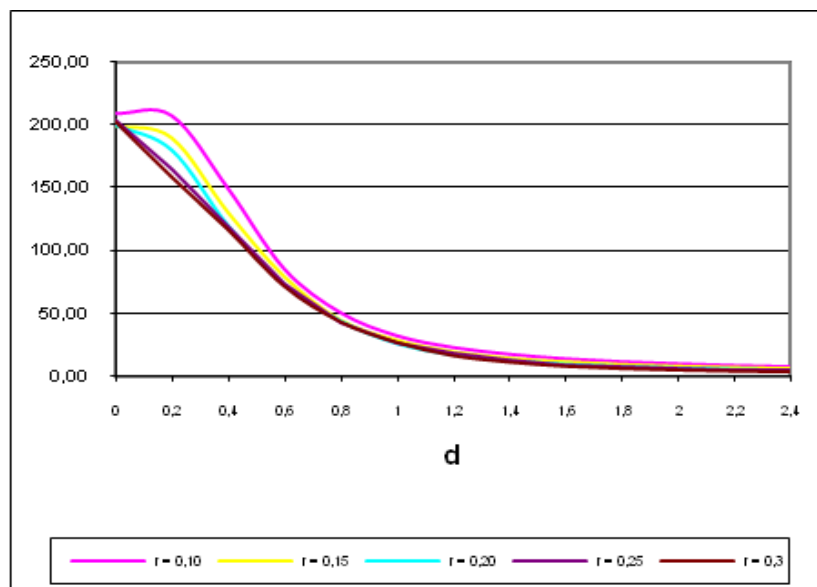


Figura 2: Curvas de ARL para $r \geq 0.10$



Para esta aplicación en particular, el análisis de lo obtenido llevó a fijar $r = 0,30$ y $h = 53,5$, como los parámetros del gráfico MEWMA a implementar.

4. Conclusiones



En este trabajo se presenta el diseño de una estrategia *MEWMA* como una poderosa herramienta para detectar cambios en los tiempos de producción requeridos para fabricar lotes de piezas en una mediana empresa metalúrgica del Gran Rosario.

Fue necesario estudiar las propiedades del gráfico *MEWMA* debido a que las distribuciones marginales de las variables consideradas eran asimétricas, con diferentes medias y desvíos estándar. Por esta razón no podían aplicarse las Tablas basadas en límites normales. Se utilizaron procedimientos de simulación para hallar los parámetros h y r del gráfico. Los valores de r y de h recomendados difieren sustancialmente de los que hubieran surgido de darse una estructura diferente para la matriz de variancias y covariancias como la considerada en los trabajos que se mencionaron. Esto enfatiza la necesidad de estudiar cada situación en particular, a fin de garantizar un adecuado comportamiento del gráfico.

Se pudo comprobar que la estrategia *MEWMA* tiene mejor performance que la del gráfico T^2 para cualquier desvío respecto del vector de medias bajo control, siendo por lo tanto su empleo más recomendable.

REFERENCIAS BIBLIOGRÁFICAS

BARBIERO, C.; FLURY, M.I.; QUAGLINO, M. RUGGIERI, M. (2006) "Gráficos Multivariados de Promedios Móviles Ponderados Exponencialmente (MEWMA) para observaciones Gamma multivariadas". Jornadas Internacionales de Estadística, Rosario.

FLURY, M.I. (2010). "Control Estadístico de Procesos con Gráficos Multivariados de Promedios Móviles Ponderados Exponencialmente (MEWMA)"- Tesis de Maestría en Estadística Aplicada- Universidad nacional de Rosario- FCEE.

KRISHNAIA, H. P. (1985). "Multivariate Gamma Distributions" in Encyclopedia of Statistical Sciences. Ed.Wiley.

LOWRY, C.; WOODALL, W.; CHAMP. C.; RIGDON, S. (1992). "A Multivariate Exponentially Weighted Moving Average Control Chart". Technometrics 34, pp.1-29.

QUAGLINO, M.; PAGURA, J.DIANDA, D.; LUPACHINI, E. (2006). "Métodos Estadísticos Aplicados para la Mejora de Procesos. Experiencia en una PYME del Gran Rosario". XI Jornadas de Investigación en la FCEyE, Rosario.

STOUMBOS, Z.; SULLIVAN, J. (2002). "Robustness to non-normality of the multivariate EWMA control chart". Journal of Quality Technology. Vol.34, No.3.



TESTIK, M.; RUNGER, G. and BORROR, C. (2003) "Robustness properties of Multivariate EWMA Control Chart". *Quality and Reliability Engineering International*, 19, pp.31-38.