

Cartas de control Bayesianas para atributos y el tamaño de subgrupo grande en la carta p

Bayesian Control Charts for Attributes and the Large Subgroup Size in p Chart

HUMBERTO GUTIÉRREZ^a

UNIVERSIDAD DE GUADALAJARA, JALISCO, MÉXICO

Resumen

Se analizan los problemas de las cartas de control para atributos. En particular, se ve que el procedimiento tradicional para obtener los límites de control tiene varias dificultades: no incorpora la incertidumbre sobre la estimación del parámetro del modelo, no contempla las variaciones en el parámetro del proceso y requiere un período base para obtener datos. Durante este período base no se tiene una carta de control para monitorear el proceso. Se ve que estos problemas son resueltos con métodos Bayesianos. Se ve cómo establecer los límites de control Bayesianos para las cartas de control p , np , c y u , en forma secuencial desde la primera observación. Esto elimina la necesidad del período base. Por último se analiza el problema de la carta p cuando esta se basa en un tamaño de subgrupo grande y se propone una solución Bayesiana para este problema.

Palabras clave: control de calidad, métodos Bayesianos, distribución beta-binomial, distribución Poisson-gama.

Abstract

The problems of the control charts for attributes are analyzed. In particular, it is seen that the traditional procedure to obtain the control limits has several difficulties: it does not incorporate the uncertainty in the estimate of the parameter of the distribution, it does not include the variation of the process, and the procedure requires a base period to obtain the control limits. Along this period it does not have a control chart to analyze the process. It is seen that these problems are solved applying Bayesian methods. It is described a sequential procedure to obtain the Bayesian control limits for p , np , c and u charts, from the first observation. This eliminates the necessity of the base period. Lastly it is analyzed the problem of the p chart when it

^aProfesor. E-mail: humpulido@yahoo.com

is based on a large subgroup size and it is proposed a Bayesian solution for this problem.

Key words: Quality control, Bayesian methods, Beta-binomial distribution, Poisson-gamma distribution.

1. Introducción

Las innovaciones crecientes en la ciencia y la técnica, y la necesidad de mejorar los niveles de productividad y competitividad en las empresas y organizaciones de todo el mundo, hacen que cada día se requieran más y mejores metodologías estadísticas que ayuden a entender (modelar) la variabilidad y que sean la base para tomar decisiones óptimas. En este contexto, los métodos de la estadística Bayesiana se están convirtiendo en una herramienta estándar en muchas aplicaciones estadísticas. En particular en las áreas de aplicación de la estadística industrial existen muchos problemas que requieren esquemas óptimos de decisión; ver por ejemplo Gutiérrez & Camacho (1996).

Un área donde se manifiesta con mucha claridad la necesidad de aplicar metodologías Bayesianas son las tradicionales cartas de control para atributos, como veremos enseguida.

En adelante con $f(y|\theta)$ denotaremos el modelo para las observaciones y , con θ el parámetro del modelo. Como es bien conocido las cartas de control ayudan a monitorear un proceso analizando la distribución de las características de calidad a través del tiempo.

En el enfoque tradicional se supone que el parámetro θ es constante, y durante un período base (PB) de observación se obtiene $\mathbf{Y} = (y_1, \dots, y_{k-1})$ para estimar θ . Con esta estimación se obtienen los límites de control y la carta se utiliza para monitorear la estabilidad del proceso a partir de la observación y_k . Por lo tanto durante el PB no se tienen los límites de control, y en consecuencia no es posible aplicar la carta de control. Esto en general es una desventaja importante para el adecuado y oportuno control de procesos, pero es sumamente crítico en los procesos en los que se obtienen datos lentamente o los procesos actuales de múltiples modelos de productos, en los cuales se trabaja cada día más con corridas cortas (Del Castillo et al. 1996, Tsiamyrtzis & Hawkins 2005).

En estos casos es casi imposible cubrir ese PB para poder calcular los límites y tener una carta de control. Otro problema que tienen todas las cartas de control tradicionales, y que en cierto sentido está ligado al PB , es que en los límites de control no se incorpora la incertidumbre sobre la estimación de θ . Formalmente se procede de igual manera si $k = 10$ o si $k = 100$.

Esta es la razón por la que se exige el período base, para de alguna forma tener información suficiente para estimar “bien” a θ . Sin embargo, como veremos más adelante, los métodos Bayesianos sí incluyen la incertidumbre con la que se estima θ y con ello no hay necesidad del PB .

Además los modelos utilizados en las cartas para atributos, como el binomial y el de Poisson, solo consideran variación del muestreo y no contemplan las pequeñas variaciones que en forma natural se dan a través del tiempo en muchos procesos; ver por ejemplo Bayarri & García-Donato (2005). Así, al no modelar adecuadamente la variación natural de los procesos, la carta de control correspondiente no será un buen instrumento para el control y monitoreo del proceso. Otro problema es que los límites de control para las cartas de atributos se obtienen con el enfoque 3-sigma, que se basa en el hecho que bajo normalidad y estabilidad, la probabilidad de que los datos estén dentro de los límites de control es de 0.9973. Sin embargo, en el caso de los datos de atributos las distribuciones son sesgadas, y en consecuencia los correspondientes límites de control no reúnen tales requerimientos (Gutiérrez & Camacho 1998, Hamada 2002). Esta problemática se acrecienta conforme los niveles de defectos son pequeños. Algo que cada día ocurre más.

En este contexto, un objetivo de este artículo es analizar los problemas anteriores y ver la manera en que las cartas de control Bayesianas para atributos los resuelven. Adicionalmente analizaremos un problema particular de las cartas p y np , que se presenta cuando el tamaño de subgrupo n es grande. Ya que en este caso los límites de control son tan estrechos que la correspondiente carta de control p o np no es útil para analizar el desempeño del proceso (Heimann 1996). Para este problema proponemos una solución con base en ideas Bayesianas.

En la sección 2 se analiza la problemática del modelado de la variación en procesos de atributos y cómo es resuelta desde la perspectiva Bayesiana. En la sección 3 se ven los detalles de la carta Bayesiana u y c y en la sección 4 se abordan las cartas de control Bayesianas p y np . El problema tamaño de subgrupo grande de la carta p se analiza en la sección 5, y las conclusiones del trabajo se presentan en la sección 6.

2. La variación de un proceso

Un aspecto importante en muchas metodologías de control de calidad para procesos de atributos es la probabilidad de que un artículo sea defectuoso, p . Por claridad en la exposición vamos a suponer que el producto ha sido fabricado por un proceso en una serie de lotes. En caso que el proceso no se produzca por lotes, entonces la producción de cada cierto período (turno, por ejemplo) se verá como un lote. Debido a fluctuaciones aleatorias, estos lotes diferirán en calidad aunque el proceso sea estable y esté en control estadístico. De cada lote se toma una muestra aleatoria y se obtiene el número de artículos defectuosos y , obviamente, y varía de muestra a muestra aunque p se mantenga constante. Las variaciones en y pueden separarse en *variación dentro de lote o debida al muestreo*, que se da porque se toma una muestra finita del lote (variación dentro de lote), y *variación entre lotes*, que se debe al muestreo y al mismo proceso. Si estas dos fuentes de variación son iguales, cada lote puede considerarse una muestra al azar obtenida de un proceso con p constante. Ésta es la premisa detrás de las cartas de control p y np (Heimann 1996) y otras metodologías tradicionales, como el muestreo de aceptación (Calvin 1984).

2.1. Variación muestral y variación del proceso

Es frecuente que la variación *entre lotes* sea mayor que la variación *dentro de lote* debido a cambios (relativamente pequeños) que se dan en el proceso a través del tiempo, por lo que p varía continuamente (Heimann 1996). Sin embargo, las metodologías estadísticas tradicionales (frecuentistas) no toman en cuenta esto, ya que consideran que p permanece constante.

Por ejemplo, en la carta de control p la variable que se lleva a la carta es la proporción $p_i = y_i/n$ de artículos defectuosos, donde y_i es el número de artículos defectuosos en un subgrupo o muestra de tamaño n . Se supone como correcto el modelo binomial, que da la probabilidad condicional de ver y defectuosos en una muestra aleatoria de tamaño n , dado el valor de p , es decir:

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, 2, \dots, n \quad (1)$$

Los límites de control 3-sigma ($E(p_i) \pm 3\sigma_{p_i}$) para la carta p se obtienen solo suponiendo variación debida al muestreo, y están dados por:

$$p \pm 3\sqrt{\frac{p(1-p)}{n}}$$

Por el contrario, desde una perspectiva Bayesiana se consideran ambas fuentes de variación, ya que la variación muestral se considera a través de la verosimilitud, y la incertidumbre sobre el valor de p se considera a través de una distribución a priori. Es usual incorporar la incertidumbre sobre el valor de p a través de la distribución a priori beta(α, β):

$$\pi(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}, \quad \alpha, \beta > 0 \text{ y } 0 \leq p \leq 1$$

Cuya media y varianza están dadas por:

$$E(p) = \frac{\alpha}{\alpha + \beta} \quad \text{y} \quad V(p) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Si denotamos $E(p) = \bar{p}$, entonces \bar{p} se puede ver como la probabilidad promedio de producir artículos defectuosos. Dado que $\bar{p} = \alpha / (\alpha + \beta)$, es fácil ver que

$$V(p) = \frac{\bar{p}(1-\bar{p})}{\alpha + \beta + 1} \quad (2)$$

De esta manera la probabilidad de obtener y artículos defectuosos en una muestra de tamaño n , no se obtiene con (1), sino con la distribución a priori predictiva (distribución marginal de y) (Bernardo & Smith 1994):

$$f(y) = \int_0^1 f(y|p)\pi(p) dp = \binom{n}{y} \frac{\Gamma(\alpha + y)\Gamma(\beta + n - y)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + n)} \quad (3)$$

Esta distribución recibe el nombre de beta-binomial o también se conoce como distribución de Polya.

Una forma de cuantificar qué tan necesario es considerar un esquema Bayesiano desde la perspectiva de los dos tipos de variación, es contrastar la variación dentro de lote contra la variación entre lotes. Sobre esto, ya dijimos que ambos tipos de variación están contemplados por la distribución beta-binomial, cuya media y varianza están dadas por:

$$E(y) = n \frac{\alpha}{\alpha + \beta} \quad y \quad V(y) = \frac{n\alpha\beta}{(\alpha + \beta)^2} \frac{(\alpha + \beta + n)}{(\alpha + \beta + 1)}$$

Luego la media y la varianza (a priori) de y/n están dadas por:

$$E(y/n) = \frac{\alpha}{\alpha + \beta} \tag{4}$$

o

$$V(y/n) = \frac{\alpha\beta}{(\alpha + \beta)^2 n} \frac{(\alpha + \beta + n)}{(\alpha + \beta + 1)}$$

Si expresamos $V(y/n)$ en términos de \bar{p} , obtenemos que

$$\begin{aligned} V(y/n) &= \frac{\bar{p}(1 - \bar{p})}{n} \frac{(\alpha + \beta + n)}{(\alpha + \beta + 1)} \\ &= \frac{\bar{p}(1 - \bar{p})}{n} \frac{(\alpha + \beta)}{(\alpha + \beta + 1)} + \frac{\bar{p}(1 - \bar{p})}{(\alpha + \beta + 1)} \end{aligned}$$

De aquí se puede apreciar que para valores moderadamente grandes de α y β , $(\alpha + \beta)/(\alpha + \beta + 1)$ va a tender a 1, y en consecuencia

$$V(y/n) \approx \frac{\bar{p}(1 - \bar{p})}{n} + \frac{\bar{p}(1 - \bar{p})}{(\alpha + \beta + 1)} \tag{5}$$

De acuerdo con las propiedades del modelo binomial y con la expresión (2), esta última expresión para $V(y/n)$ (5) se puede ver como una suma de la variación de un proceso binomial más el de la distribución beta. Es decir, igual a la variación del muestreo más la variación del proceso (en términos de \bar{p}). De aquí resulta razonable obtener la razón de varianza total con la varianza del muestreo (binomial). Es fácil ver que está dada por:

$$r = \frac{V(y/n)}{\frac{\bar{p}(1 - \bar{p})}{n}} = \frac{\alpha + \beta + n}{\alpha + \beta + 1} \tag{6}$$

Así, valores de r mayores que 1 significan que la variación observada no sólo se debe a la variación del muestreo (como se supone tradicionalmente), sino que además está presente la variación debido a que p no se mantiene constante en el proceso a través del tiempo.

Siguiendo ideas de cartas de control, Heimann (1996) establece que el valor máximo de r donde todavía el modelo binomial puede ser adecuado es de 1.357.

De esta manera el parámetro r se convierte en un indicador para decidir cuándo usar un esquema Bayesiano. El recíproco de r , $1/r$, es un indicador de la proporción de la variación total que se debe al muestreo.

Si en los procesos tradicionales es difícil que p se mantenga constante, en los procesos de hoy en día es mucho menos probable que p permanezca constante. Ya se tienen múltiples modelos del mismo producto, los tamaños de lote son cada día más pequeños debido a la exigencia de no producir para inventario. Por otra parte, cada día más los productos son resultados de ensambles o componentes, que a su vez se produjeron en distintas plantas ubicadas en diferentes regiones o países. Todo esto hace que se tengan mayores fuentes de variación a través del tiempo (lote a lote) y que en consecuencia p no permanezca constante. Otros trabajos se han originado a partir de que p no permanece constante, ni siquiera dentro del lote (Quarshie & Shindo 1996).

2.2. Información a priori

Tanto para aplicar la distribución beta-binomial como para estimar el parámetro r , ver (6), es necesario especificar los hiperparámetros de la distribución beta(α, β). Para estimar r básicamente se ha propuesto obtener la variación de las proporciones en m muestras previas y compararla con la variación bajo el modelo binomial, de la siguiente manera:

$$r = \frac{\frac{1}{m-1} \sum_{i=1}^m (p_i - \tilde{p})}{\tilde{p}(1 - \tilde{p})/n} \quad (7)$$

donde p_i es la proporción de defectuosos en la muestra i , y \tilde{p} es la proporción de defectuosos obtenidas en las m muestras (Calvin 1984, Heimann 1996).

De acuerdo con lo anterior, para obtener los valores de α y β , veamos que (4) y (6) son dos ecuaciones simultáneas que tienen como incógnitas a α y β . Resolviendo estas ecuaciones se obtiene que:

$$\alpha = \bar{p} \left(\frac{n-r}{r-1} \right) \quad \text{y} \quad \beta = (1 - \bar{p}) \left(\frac{n-r}{r-1} \right)$$

donde el valor de \bar{p} se puede estimar con \tilde{p} .

En el caso más realista en el que no se disponga de las m muestras iniciales, para determinar α y β , y con ello r , se puede utilizar el procedimiento que se propone en Gutiérrez & Aguirre (2005), donde es necesario que a priori se dé un intervalo en el que se espera esté p . Sea $[L_p, U_p]$ este intervalo.

Aplicando el procedimiento propuesto en el trabajo referido, los valores de α y β están dados por:

$$\alpha = \frac{z^2 (2 - L_p - U_p) (L_p + U_p)^2 - (U_p - L_p)^2 (L_p + U_p)}{2 (U_p - L_p)^2} \quad (8)$$

y

$$\beta = \alpha \left(\frac{2 - L_p - U_p}{L_p + U_p} \right) \tag{9}$$

donde z es un número que se recomienda esté entre $[1.5, 3.5]$. Este valor de z está directamente relacionado con la probabilidad inicial de que $[L_p, U_p]$ abarque la variación de p . A mayores valores de z más probabilidad y en consecuencia una mayor porción de la densidad a priori $\text{beta}(\alpha, \beta)$ estará en $[L_p, U_p]$. Una forma sencilla de elegir z es verlo como si fuera el cuantil de una distribución normal estándar, donde se conoce la porción de la densidad que está entre $-z$ y z . De esta manera, si se cree que el 95% de la distribución a priori está entre L_p y U_p , entonces se puede elegir un valor de z cercano a 2.

2.3. Efecto de no considerar la variación de p

En este apartado veremos un ejemplo del tipo de efecto que tiene no considerar la variación en p . Lo haremos a través de la obtención de una curva característica de operación (CO) de un plan de muestreo de aceptación (Gutiérrez & De la Vara 2004, cap. 12). Si las características de calidad son variables de atributos del tipo pasa-no-pasa, entonces un plan simple de MA del tipo (N, n, c) está definido por un tamaño de lote N , un tamaño de muestra n , y el número de aceptación c . Por ejemplo el plan $N = 6000, n = 200$ y $c = 2$; significa que de un lote de 6000 unidades se seleccionan e inspeccionan 200; y si el número de artículos defectuosos en la muestra y es menor o igual que c ($y \leq c$), entonces el lote es aceptado. Pero si $y > c$, el lote es rechazado. Si se tiene un escenario donde se quiere un plan para atributos del tipo (N, n, c) , con $(n/N) < 0.10$, y p permanece constante, entonces estamos bajo el caso del muestreo de aceptación tradicional donde la probabilidad de aceptación (P_a) de un lote está dada por el modelo binomial (1):

$$P_a = \sum_{y=0}^c f(y|p) \tag{10}$$

Si por el contrario p no permanece constante en el proceso, entonces estamos bajo el escenario de un muestreo de aceptación Bayesiano, donde P_a está dada por la distribución beta-binomial (3):

$$P_a = \sum_{y=0}^c f(y) \tag{11}$$

Ejemplo 1. Supongamos un plan con $N = 3500, n = 200, c = 3$. Para calcular la CO para el muestreo de aceptación tradicional simplemente se aplica (10) utilizando algún software computacional apropiado.

En el caso Bayesiano, supongamos que para un valor dado de p , este puede variar hasta en un $\pm 0.5p$. Con este criterio se obtiene el intervalo $[L_p, U_p]$ para cada p en la curva CO, y con (8) y (9) se obtienen α y β .

En la figura 1 se muestran las curvas CO para el muestreo tradicional que no incluye variación en p , y para el Bayesiano que sí lo considera. Como se aprecia,

el MA tradicional subestima la probabilidad de aceptar el lote conforme se incrementa p . Por ejemplo para valores de p cercanos a 0.032, las probabilidades de aceptar son 0.118 y 0.174, para los casos tradicional y Bayesiano, respectivamente. Por lo tanto, si hay variaciones de p , entonces si se aplica un plan tradicional del tipo (N, n, c) , se estará dejando pasar peores niveles de calidad de los que muestra la correspondiente curva CO. Por ello en estos casos será mejor cuantificar la variación de p con una distribución a priori y diseñar el plan de muestreo Bayesiano con base en (11).

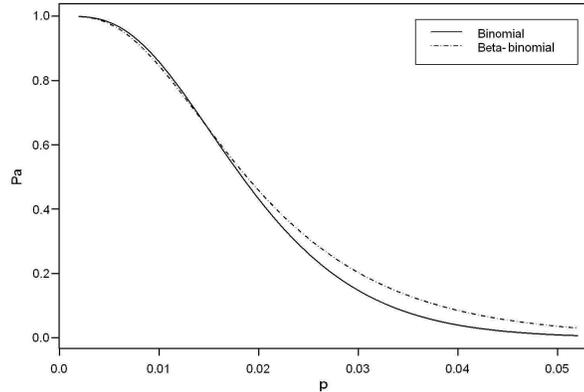


FIGURA 1: Curvas CO para MA tradicional y Bayesiano para el plan $n = 200$ y $c = 3$.

3. Cartas Bayesianas u y c

En las tradicionales cartas de control u , la característica de calidad es el número de defectos por unidad de inspección. Así, en esta carta se analiza la variable $u = y/m$, donde y es el número de defectos en una unidad de inspección de tamaño m . Cuando la unidad de inspección es constante esta carta se reduce a la carta c (Gutiérrez & De la Vara 2004, cap. 8).

Los límites de control para estas cartas se derivan bajo el supuesto de que $f(y|\lambda) \sim Poisson(m\lambda)$:

$$f(y|\lambda) = \frac{e^{-m\lambda}(m\lambda)^y}{y!}$$

con $\lambda > 0$ y $y = 0, 1, 2, \dots$

En forma tradicional el parámetro λ es estimado con el usual estimador insesgado:

$$\bar{u} = \frac{\sum_{i=1}^k y_i}{\sum_{i=1}^k m_i} \quad (12)$$

donde m_i es el tamaño del conjunto (subgrupo) del cual se obtuvo y_i . Los tradicionales límites de control 3-sigma para la carta u están dados por:

$$\bar{u} \pm 3\sqrt{\bar{u}/m_i} \quad (13)$$

Un primer problema con estos límites es que no se incorpora de ninguna manera el número de observaciones con base en las que se calculó \bar{u} . Es decir, no se incorpora la incertidumbre sobre la estimación de λ , y se procede igual si \bar{u} se obtuvo con $k = 5$ subgrupos que si se hizo con $k = 500$. De ahí que la recomendación tradicional de tener por lo menos 20 subgrupos durante el *PB* se vuelva crítica, ya que de esa manera se trata de tener un “buen” estimador puntual para λ . Sin embargo, en los ambientes actuales de manufactura y en los procesos lentos es cada día más difícil cumplir con tal recomendación.

Además la idea de los límites de control 3-sigma se basa en el hecho que, bajo normalidad y estabilidad, la probabilidad de que los datos estén dentro de los límites de control es de 0.9973. Sin embargo en el caso de los datos de atributos las distribuciones son sesgadas, y en consecuencia los correspondientes límites de control no reúnen tales requerimientos (Gutiérrez & Camacho 1998, Hamada 2002).

Por otro lado, algunos autores (Irony & Pereira 1994) han reportado que es frecuente que el modelo Poisson tenga un pobre ajuste para este tipo de datos. Además, al considerar que el parámetro λ permanece constante, no se considera la variación de corto plazo. Pero como ya lo vimos en la sección anterior, es frecuente que los procesos tengan variaciones de corto plazo aun en el caso de procesos con una razonable estabilidad.

Por lo anterior estamos ante el campo propicio de aplicar un modelo Bayesiano para cartas de control. En el caso particular de las cartas u y c , se supone que λ sigue una distribución a priori gama(α, β) y por lo tanto la probabilidad de encontrar y defectos en una unidad de inspección de tamaño m , la da la distribución marginal (a priori) predictiva, que se puede demostrar que está dada por:

$$f(y) = \int_0^{\infty} f(y|\lambda)\pi(\lambda)d\lambda = \frac{\Gamma(y + \alpha) m^y}{\Gamma(\alpha)(m + \beta)^y y!} \left(\frac{\beta}{m + \beta}\right)^\alpha \quad (14)$$

Esta distribución se conoce en la literatura Bayesiana como Poisson-gama. Para valores enteros del hiperparámetro α , esta distribución se reduce a una distribución binomial negativa ($\alpha, m(m + \beta)^{-1}$), que algunos autores la han propuesto como un mejor modelo para datos de atributos (Sheaffer & Leavenworth 1976). Así el modelo (14) está tomando en cuenta tanto la posible variación de λ que en forma natural se da en el proceso (a través de la distribución a priori), como la variación debida al muestreo (a través de la verosimilitud).

Un aspecto clave en un contexto Bayesiano es evitar el *PB*, ya que si se tiene un conocimiento genuino sobre el proceso, se puede obtener una distribución a priori para λ , y con esta se puede obtener (14) para que desde la primera observación del proceso se tenga una carta de control. Incluso en el caso que no se conozca nada y la distribución a priori sea poco informativa, con relativamente baja cantidad de observaciones se tendrá una buena carta Bayesiana u (Hamada 2002).

En Gutiérrez & Aguirre (2005) se propone un procedimiento para determinar los hiperparámetros de la distribución a priori $\text{gama}(\alpha, \beta)$. Parten de que el experto en el proceso dé un intervalo $[L_\lambda, U_\lambda]$ donde se espera que esté λ . Aspecto que es razonable suponer, debido a que λ representa el número promedio de defectos por pieza o artículo. A partir de tal intervalo, se obtiene que los hiperparámetros están dados por:

$$\beta = \frac{2z^2(L_\lambda + U_\lambda)}{(U_\lambda - L_\lambda)^2} \quad (15)$$

$$\alpha = \left(\frac{z(L_\lambda + U_\lambda)}{U_\lambda - L_\lambda} \right)^2 \quad (16)$$

donde z se asigna, como comentamos en el párrafo siguiente, a las ecuaciones (8) y (9).

Supongamos que se ha obtenido del proceso $k - 1$ observaciones: u_1, \dots, u_{k-1} ; con $u_i = y_i/m_i$. Es fácil ver que dados $Y = (y_1, \dots, y_{k-1})$, $M = (m_1, \dots, m_{k-1})$, $s = \sum_{i=1}^{k-1} y_i$ y $m_s = \sum_{i=1}^{k-1} m_i$, la distribución posterior está dada por:

$$\pi(\lambda|Y, M) = \frac{(m_s + \beta)^{s+\alpha}}{\Gamma(s + \alpha)} e^{-\lambda(m_s + \beta)} \lambda^{s+\alpha-1}$$

que corresponde a una densidad $\text{gama}(s + \alpha, m_s + \beta)$ (Bernardo & Smith 1994). A partir de aquí se obtiene la correspondiente distribución posterior predictiva para la siguiente observación en términos de y :

$$\begin{aligned} f(y_k|Y) &= \int_0^\infty f(y_k|\lambda)\pi(\lambda|Y)d\lambda \\ &= \frac{\Gamma(y_k + \alpha + s)}{\Gamma(\alpha + s)(y_k)!} \left(\frac{m_k}{\beta + m_s + m_k} \right)^{y_k} \left(\frac{\beta + m_s}{\beta + m_s + m_k} \right)^{s+\alpha} \end{aligned} \quad (17)$$

m_k es el tamaño de subgrupo para y_k . Por lo tanto los límites de control inferior (LCI) y superior (LCS) para una carta Bayesiana en términos de y están dados por:

$$YLCI_k = q_{\gamma/2} \quad \text{y} \quad YLCS_k = q_{1-\gamma/2} \quad (18)$$

con q_γ , el cuantil γ de la distribución (17). Se ha agregado el subíndice k a los límites de control para enfatizar que esos límites son el pronóstico de donde se espera la observación y_k con tamaño de subgrupo m_k , de acuerdo con lo que se ha observado hasta el momento $k - 1$.

Si se quieren los límites de control Bayesianos para la carta u , de manera predictiva para $u_k = y_k/m_k$, entonces como (17) es una distribución discreta, los límites dados por (18) simplemente se dividen por m_k :

$$ULCI_k = q_{\gamma/2}/m_k \quad \text{y} \quad ULCS_k = q_{1-\gamma/2}/m_k \quad (19)$$

Así conforme se vaya obteniendo información los límites de control se van calculando con la información disponible hasta ese momento y los límites se ven como un pronóstico sobre el desempeño futuro del proceso. La línea central se obtiene con la media de la distribución predictiva, que para el caso de la carta u está dada por $LC = (\alpha + s)/(\beta + m_s)$.

Si se observara un dato que cae fuera de los límites de control, se tendría que investigar inmediatamente si se ha dado un cambio especial en el proceso, en cuyo caso esa información no deberá integrarse a la información con la que se están actualizando los límites de control. Este proceso secuencial de recalculer los límites con cada nueva observación se puede suspender hasta que se tenga un historial suficiente, para que una vez establecidos los límites se puedan detectar derivas o tendencias en el proceso. Cuando haya un cambio mayor se puede reiniciar el cálculo de límites, partiendo de una distribución a priori.

Respecto a la carta c , como habíamos dicho esta se aplica cuando la unidad de inspección en la carta u es constante y por ello deja de intervenir en los cálculos el tamaño m de la unidad de inspección. Específicamente, en este caso (17) toma la forma siguiente:

$$f(y_k|Y) = \left(\frac{\beta + k - 1}{\beta + k}\right)^{s+\alpha} \frac{\Gamma(y_k + \alpha + s)}{\Gamma(\alpha + s)(y_k)!} \left(\frac{1}{\beta + k}\right)^{y_k}$$

Los límites de control para la carta c se obtienen con los cuantiles $\gamma/2$ y $1 - \gamma/2$ de esta distribución.

Ejemplo 2. En Gutiérrez & De la Vara (2004, cap. 8) se presentan los defectos encontrados en la inspección por muestreo de 24 lotes de piezas electrónicas. Los datos se muestran en la tabla 1.

TABLA 1: Defectos encontrados y_i en m_i piezas de 24 lotes.

Lote	1	2	3	4	5	6	7	8	9	10	11	12
m_i	20	20	20	20	15	15	15	25	25	25	25	30
y_i	17	24	16	26	15	15	20	18	26	10	25	21
Lote	13	14	15	16	17	18	19	20	21	22	23	24
m_i	30	30	30	30	30	30	15	15	15	15	15	15
y_i	40	24	46	32	30	34	11	14	30	17	18	20

Si se aplica la carta u tradicional es necesario esperar los 24 lotes para estimar \bar{u} ; de acuerdo con (17), $\bar{u} = 1.046$. Los límites de control se obtienen según (13). En la figura 2a se muestra la correspondiente carta de control u . De donde se aprecia que en los lotes 10 y 21 ocurren hechos especiales en el proceso. Pero esto lo sabemos hasta después de producido el lote 24, por lo que es demasiado tarde para actuar.

Para poder obtener la carta Bayesiana u , y obtener límites secuenciales, lo primero que necesitamos es especificar los parámetros de la distribución a priori para λ . Si suponemos que λ está entre $[0.5, 1.5]$, con una probabilidad aproximada de 0.95. De esta manera, de acuerdo con (16) y (15), y tomando $z = 2$, los parámetros para la distribución a priori gama son $\alpha = 16$ y $\beta = 16$. En la figura 2b se aprecia la correspondiente carta Bayesiana u . Para los límites de control se usa la significancia usual ($\gamma = 0.0027$) y se van obteniendo de manera secuencial de acuerdo con (19).

Contrastando la carta u tradicional (figura 2a) con la carta Bayesiana u (figura 2b), se puede apreciar lo siguiente. Con la carta Bayesiana desde la primera observación se tenían límites de control: con base en la distribución a priori se esperaba que u_1 estuviera entre 0.2 y 2.2. Estos límites son los más amplios comparados con los de la carta u tradicional. Una vez obtenida la primera observación, los límites para la segunda observación en la figura 2b, prácticamente se parecen a los límites de la carta u tradicional. Los límites de la carta Bayesiana son exactos y no del tipo 3-sigma. La carta Bayesiana detectó un punto adicional fuera del límite de control superior (la observación 15).

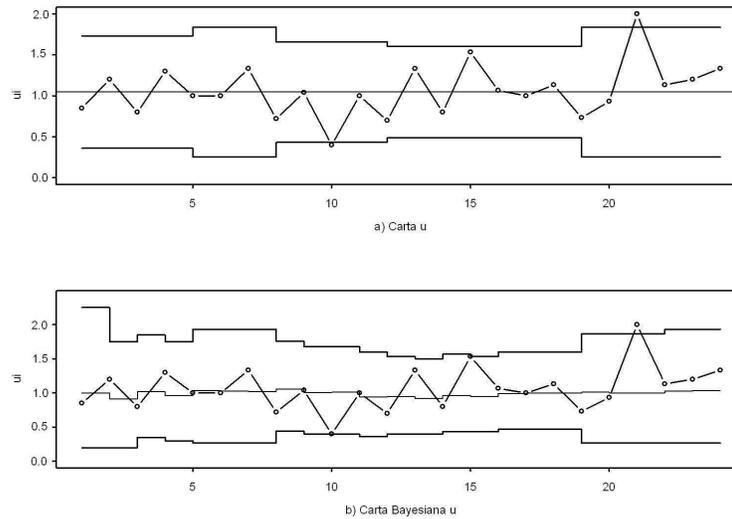


FIGURA 2: Cartas de control para datos de la tabla 1.

4. Cartas Bayesianas p y np

En la carta de control p se lleva la proporción $p_i = y_i/n_i$ de artículos defectuosos por subgrupo de tamaño n_i . En forma tradicional los límites de la carta se calculan bajo el supuesto de distribución binomial. Los límites 3-sigma para esta carta están

dados por:

$$\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} \tag{20}$$

donde \bar{p} es el estimador insesgado de p :

$$\bar{p} = \frac{\sum_{i=1}^k y_i}{\sum_{i=1}^k n_i}$$

Al igual que en el caso de la carta u , en la estimación p no se incorpora de ninguna manera el número de observaciones con base en las que se calculó \bar{p} . Es decir, no se incorpora la incertidumbre sobre la estimación de p y se procede igual si \bar{p} se obtuvo con $k = 10$ subgrupos que si se hizo con $k = 200$. También los límites 3-sigma tradicionales para la carta p tienen el problema de no dejar una probabilidad de 0.00135 en cada una de las colas de la distribución (Gutiérrez & Camacho 1998, Hamada 2002). Además se está suponiendo que p permanece constante. Pero ya hemos visto que existen variaciones de corto plazo que hacen que p no se mantenga constante, aun para procesos con una moderada estabilidad.

Los puntos anteriores junto con el hecho de que durante el PB no se pueda tener límites de control, son razón suficiente para justificar la utilización de una carta Bayesiana p , que contemple variaciones en p , y que desde la primera observación puede tener límites de control. Con distribución a priori $\text{beta}(\alpha, \beta)$ la distribución posterior para p está dada por:

$$\pi(p|Y) = \frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*-1} (1-p)^{\beta^*-1} \tag{21}$$

que corresponde a una distribución beta

$$\alpha^* = \alpha + s, \beta^* = \beta + n_s - s \quad \text{con} \quad s = \sum_{i=1}^{k-1} n_i p_i, \quad n_s = \sum_{i=1}^{k-1} n_i$$

La distribución posterior predictiva es de la forma beta-binomial y está dada por:

$$f(y_k|Y) = \binom{n_k}{y_k} \frac{\Gamma(\alpha^* + \beta^*)\Gamma(\alpha^* + y_k)\Gamma(\beta^* + n_k - y_k)}{\Gamma(\alpha^*)\Gamma(\beta^*)\Gamma(\alpha^* + \beta^* + n_k)} \tag{22}$$

De esta manera los límites de control para la carta Bayesiana p están dados por:

$$LCI_k = q_{\gamma/2}/n_k \quad \text{y} \quad LCS_k = q_{1-\gamma/2}/n_k, \tag{23}$$

donde q_γ es el cuantil γ de la distribución (22). Obviamente estos límites estarían lo más próximo posible (por lo discreto de la distribución) a la cobertura $1 - \gamma$ deseada, y en consecuencia no tienen el problema de cobertura de los límites 3-sigma. La línea central se obtiene con $LC = \alpha^*/(\alpha^* + \beta^*)$. En la siguiente sección veremos un ejemplo de carta Bayesiana p (ver figura 3d).

Cuando el tamaño de subgrupo es constante se suele utilizar la carta np , donde se lleva en la carta directamente el número de defectos por subgrupo. En este caso los límites de control están dados directamente por los cuantiles $\gamma/2$ y $1 - \gamma/2$ de la densidad (22).

5. Carta Bayesiana p y el tamaño de subgrupo grande

Un hecho relevante en el cálculo de los límites de control en (20) es que para tamaños de subgrupo n_i grandes la amplitud de los correspondientes límites de control se vuelven sumamente estrechos. De tal forma que proporciones de subgrupo con pequeñas desviaciones con respecto a \bar{p} caen fuera de los límites. Esta estrechez puede ser tal que existen casos en los cuales casi ningún punto cae dentro de los límites (Gutiérrez & Camacho 1996, Heimann 1996). Evidentemente en estos casos la carta p resulta de nula utilidad.

Como alternativa al anterior problema, Gutiérrez & Camacho (1996) y Heimann (1996) proponen el uso de carta de individuales, haciendo $X_i = p_i$. Es bien conocido que los límites de una carta de individuales están dados por:

$$\bar{X} \pm 3\bar{R}/1.128 \quad (24)$$

En el caso de las proporciones $\bar{X} = \bar{p}$, y \bar{R} es el promedio de los rangos móviles de orden 2. De esta manera en los límites obtenidos con (24) ya no influye n . En Gutiérrez & Camacho (1996) se demuestra mediante un estudio de simulación que si la carta p no es apropiada cuando n es grande, es porque hay una sobredispersión en p_i no contemplada por el modelo binomial. De esto ya se habló al inicio de la sección anterior.

En Heimann (1996) utilizando el estadístico (7) se llega a la conclusión que el valor máximo de n que sería congruente con el modelo binomial está dado por:

$$n^* = 1.357\bar{p}/\sigma_T^2 \quad (25)$$

donde $\sigma_T^2 = \frac{1}{k-1} \sum_{i=1}^k (p_i - \bar{p})$. De esta manera si en un caso particular el n utilizado es mayor que n^* , entonces la tradicional carta p no es adecuada porque hay una sobredispersión y los límites de control serán muy estrechos.

Ejemplo 3. En Heimann (1996) se describe un problema de control de la calidad del servicio telefónico, donde se registra semanalmente el número de circuitos que al repararlos se excede un tiempo previsto, y lo dividen entre el total de circuitos reparados. El tamaño promedio de circuitos reparados es de $n = 7700$. En la figura 3a se muestra la tradicional carta de control p con límites obtenidos de acuerdo con (20). A partir de esta gráfica es evidente que la carta p es inapropiada para analizar el estado de tal proceso, debido a que los límites de control son demasiado estrechos. Esto indica que el proceso tiene una sobredispersión no contemplada por

el modelo binomial. Esto lo refleja el estadístico r , ver (6), que en este caso es igual a 25.3, muy superior al 1.357 que se considera como aceptable para considerar que la variación solo se debe al muestreo.

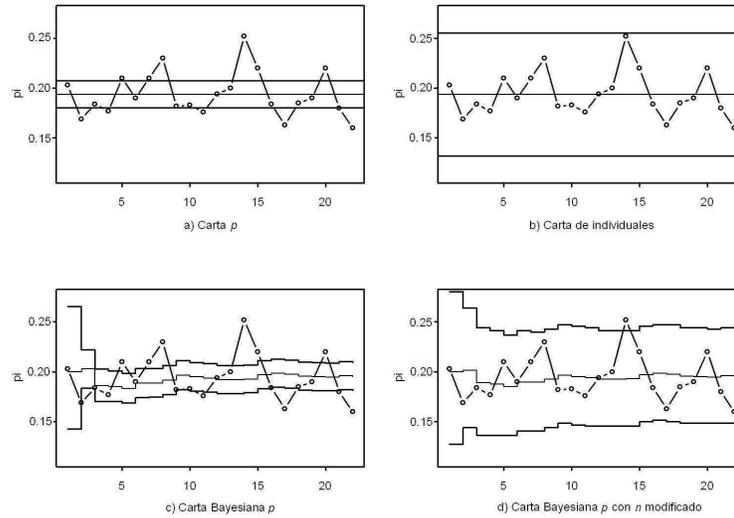


FIGURA 3: Cartas de control con n grande para reparaciones fuera de tiempo.

En la figura 3b se analizan las mismas proporciones con una carta de individuales, donde los límites de control son $LCI = 0.132$ y $LCS = 0.255$, que se han obtenido con (24). Esta gráfica refleja mejor el estado del proceso, y con ella se pueden detectar cambios grandes en su desempeño. Sin embargo tanto en este caso como en otros presentados en Heimann (1996) y Gutiérrez & Camacho (1996) se aprecia que los límites de control de la carta de individuales para proporciones resultan demasiado amplios, de tal forma que solo permiten detectar cambios grandes en el proceso. Una alternativa, que en ocasiones se recomienda en las cartas de individuales, es utilizar límites 2-sigma. En este caso estos límites están dados por 0.152 y 0.244, que se ajustan mejor a la variación del proceso. Pero obviamente esto no sería una forma ortodoxa de proceder.

Por otro lado, ya vimos que la carta Bayesiana p corrige varios de los problemas que tiene la carta p tradicional. Veamos qué pasa en el caso de n grande. Utilizando como información a priori para p el intervalo $[L_p, U_p] = [0.16, 0.24]$, y aplicando (8) y (9), con $z = 2$, se obtienen los parámetros α y β de la distribución a priori beta(79.8, 319.2), y de manera secuencial se obtienen los límites de control. En la figura 3c se muestra la carta Bayesiana p . Es claro que estos límites de control, obtenidos con (23), son también demasiado estrechos. Estudiamos otros casos y vimos que en general con n muy grande esto tiende a presentarse.

La explicación es sencilla: aunque en la carta Bayesiana p se permite la variación en p a través de la distribución a priori, se sigue utilizando un modelo binomial y cuando n es muy grande la amplitud de la verosimilitud correspondiente es muy

estrecha y también es muy dominante sobre la distribución a priori. Todo esto se recoge a través de los parámetros de la distribución posterior; ver (21).

Por ello proponemos como alternativa utilizar el estadístico r , ver (6), que proporciona información sobre qué tanta dispersión no es considerada por el modelo binomial. En el ejemplo $r = 25.3$, y por lo tanto el modelo binomial solo explica el 3.95 % de la variabilidad observada ($1/r$). De acuerdo con (25), el valor máximo de n que sería congruente con un modelo binomial sería $n^* = 513$. De esta manera si utilizamos este como el valor de n en (20), entonces los límites de control de la carta p para los datos bajo análisis serían: $LCI = 0.141$ y $LCS = 0.246$. Estos límites son ligeramente más estrechos que los de la carta de individuales, y por lo dicho parecen reflejar mejor la variación observada en la figura 3a. Así que si se quiere utilizar la tradicional carta p , y n es grande, una primera alternativa que proponemos es utilizar como valor de n el dado por (25).

Sin embargo, si queremos aprovechar las ventajas de una carta Bayesiana p y detectar mejor cambios moderados en el proceso, entonces proponemos que el valor de n considerado para obtener los parámetros de la distribución posterior esté dado por wn^* , donde de acuerdo con las evaluaciones que hemos realizado el valor de $w \in [1.0, 2.0]$. Conforme w sea más grande los límites de control serán más estrechos. Por ejemplo, en el caso que nos ocupa, utilizando la misma información a priori para p que se utilizó en la carta Bayesiana p , y haciendo $w = 1.3$, se toma $n_i = 1.3 \times 513 = 667$ para obtener los parámetros de la distribución posterior de manera secuencial; ver (21). La carta Bayesiana p con $n = 667$ se muestra en la figura 3d. Es claro que esta carta refleja mejor la variación del proceso, y por tanto con ella se puede detectar mejor sus cambios. Con esta carta se ha detectado que en la semana 14 el proceso tuvo un desempeño peor de lo usual. Además desde la segunda observación (subgrupo) se tienen límites de control útiles para monitorear el proceso, y no es necesario esperar a tener 20 observaciones como tiene que hacerse en la carta p tradicional.

6. Conclusiones

Hemos visto que el uso de las cartas de control Bayesianas para atributos corrige varios de los problemas de las tradicionales cartas de atributos, como: no incorporar la incertidumbre con la que se estima el parámetro del modelo, no permitir variaciones pequeñas en tal parámetro, límites 3-sigma que no tienen la cobertura buscada en las cartas de control, y recurrentes períodos base durante los que no se dispone de límites en las cartas de control, lo que impide monitorear adecuadamente el desempeño del proceso durante tales períodos base.

Si en general las cartas de control Bayesianas tienen ventajas sobre las tradicionales, en el caso de procesos lentos o de corridas cortas estas ventajas se incrementan debido a que a partir de la información a priori se pueden tener límites de control desde el inicio y conforme se van obteniendo nuevos datos los límites se van recalculando en forma secuencial durante un período de tiempo (ver párrafo

antes del ejemplo 2). Lo anterior se facilita con la forma práctica que hemos incluido para especificar los parámetros de las distribuciones a priori para los modelos binomial y de Poisson.

Se ha visto que la carta p no funciona con tamaños de subgrupo grandes. Hemos propuesto utilizar la información respecto a la sobredispersión en las proporciones muestrales que no considera el modelo binomial y con base en eso utilizar una carta Bayesiana p con tamaño de muestra modificado, dado por $n = wn^*$, con n^* dado por (25) y $w \in [1.0, 2.0]$.

Recibido: septiembre de 2006

Aceptado: octubre de 2006

Referencias

- Bayarri, M. & García-Donato, G. (2005), 'Bayesian Sequential Look at u -control Charts', *Technometrics* **47**(2), 142–151.
- Bernardo, J. & Smith, A. (1994), *Bayesian Theory*, Wiley & Sons, New York.
- Calvin, T. (1984), How and When to Perform Bayesian Acceptance Sampling, in '38th Annual Quality Congress', American Society for Quality, Milwaukee, EU, pp. 141–145.
- Del Castillo, E., Grayson, J., Montgomery, D. & Runger, G. (1996), 'A Review of Statistical Process Control Techniques for Short Run Manufacturing Systems', *Communications in Statistics-Theory and Methods* **25**(11), 2723–2737.
- Gutiérrez, H. & Aguirre, V. (2005), A Method of Moments Procedure for Eliciting Prior Distributions, Reporte técnico de-c05.8, Instituto Tecnológico Autónomo de México (ITAM), México.
- Gutiérrez, H. & Camacho, O. (1996), Ineficiencia de la carta p para tamaños de subgrupo grandes: diagnóstico y alternativas., in 'Memorias del X Foro Nacional de Estadística y del II Congreso Iberoamericano de Estadística', Asociación Mexicana de Estadística e Inegi, México.
- Gutiérrez, H. & Camacho, O. (1998), 'Modificación de las cartas de atributos (p , np , cyu) para mejorar su aproximación a la significancia bajo normalidad', *Agrociencia* **32**(4), 385–394.
- Gutiérrez, H. & De la Vara, R. (2004), *Control estadístico de calidad y seis sigma*, McGraw-Hill, México.
- Hamada, M. (2002), 'Bayesian Tolerance Interval Control Limits for Attributes', *Quality and Reliability Engineering International* **58**(1), 45–52.
- Heimann, P. (1996), 'Attributes Control Charts with Large Sample Sizes', *Journal of Quality Technology* **28**(4), 451–459.

- Irony, T. & Pereira, C. (1994), 'Motivation for the Use of Discrete Distribution in Quality Assurance', *Test* **3**, 181–193.
- Quarshie, B. & Shindo, H. (1996), 'A Comparison of Operating Characteristics of the $p - V$ and the $p - Rp$ Charts', *Quality Engineering* **9**(2), 221–228.
- Sheaffer, R. & Leavenworth, R. (1976), 'The Negative Binomial Model for Counts in Units of Varying Size', *Journal of Quality Technology* **8**, 158–163.
- Tsiamyrztis, P. & Hawkins, D. (2005), 'A Bayesian Scheme to Detect Changes in the Mean of a Short-Run Process', *Technometrics* **47**(4), 446–456.