

ANÁLISIS DE LOS EFECTOS DE LA EXISTENCIA DE AUTOCORRELACIÓN SOBRE LA GRÁFICA C DE CONTROL DE ATRIBUTOS UTILIZANDO EL MODELO INAR(1)

SERGIO M. BUZZI - MARA LLOP - ANDREA F. RIGUETTI - SILVIA JOEKES
Facultad de Ciencias Económicas-Universidad Nacional de Córdoba - ARGENTINA
sergio.buzzi@eco.unc.edu.ar- mara_llop@hotmail.com
analizamos@yahoo.com.ar - joekess@eco.unc.edu.ar

Fecha Recepción: Noviembre 2012 - Fecha Aceptación: Julio 2013

RESUMEN

El objetivo central del presente trabajo es analizar las consecuencias que tiene la existencia de correlación en el gráfico "c" de control de atributos. Para esto se adopta la estrategia de generar datos con distribución similar a una distribución Poisson pero introduciendo autocorrelación, lo que se realiza empleando un modelo autorregresivo entero de orden uno (INAR(1)). Luego se utilizan los datos simulados para analizar el desempeño de la gráfica citada, encontrándose que la capacidad de la misma para detectar cambios en la media del proceso no se ve afectada.

Por último, se propone una regla de rachas para detectar salidas de control del proceso debidas a cambios en la autocorrelación y se analiza su desempeño utilizando técnicas de simulación.

PALABRAS CLAVE: Control Estadístico de Procesos - Gráfico c – Series de Tiempo – Autocorrelación – Modelo INAR(1).

ABSTRACT

The central aim of this paper is to analyze the implications of the existence of autocorrelation in the c-chart for attributes control. In order to this, the strategy adopted is to generate data with similar distribution to Poisson distribution but introducing autocorrelation, which is performed using an integer autoregressive model of order one (INAR (1)). Then, the simulated data is used to analyze the performance of the above-mentioned graph, finding that its capacity to detect changes in the process mean is not affected.

Finally, we propose a runs rule to detect out of control proceses due to changes in the autocorrelation and we analyze its performance using simulation techniques.

KEYWORDS: Statistical Process Control - C-chart – Time Series – Autocorrelation – INAR(1) model.

1 INTRODUCCIÓN

En el caso del monitoreo de variables continuas autocorrelacionadas, una de las metodologías más utilizadas para construir gráficos de control es el empleo de modelos autorregresivos (AR). Por ejemplo puede consultarse ALWAN y ROBERTS (1988). En particular en el caso del modelo AR(1), se tiene:

$$x_t = \alpha x_{t-1} + \varepsilon_t$$

donde $\varepsilon_t \sim N(0, \sigma^2)$. Por otra parte los ε_t son i.i.d. (independientes e idénticamente distribuidos) y el parámetro α no supera a 1 en valor absoluto para que el proceso sea estacionario.

La introducción del término autorregresivo produce que, si bien los errores son independientes, la variable x en un determinado periodo estará correlacionada con su valor previo. Se conoce que el coeficiente de autocorrelación de este proceso es igual al coeficiente autorregresivo α .

El uso de los modelos AR no es factible en el caso de control estadístico de procesos discretos, porque los mismos están contruidos en base a errores normales y por lo tanto continuos. Por otra parte, no basta con simplemente utilizar una distribución discreta para los errores, dado que la parte autorregresiva no será discreta porque se obtiene como una fracción del valor previo de la variable.

Se han desarrollado varios modelos teniendo en mente dos objetivos: que asuman valores discretos y que tengan un comportamiento equiparable a los AR, especialmente que tengan similar estructura de autocorrelación. En particular, una de las familias de modelos más utilizadas, es la de los modelos INAR (procesos autorregresivos enteros,) desarrollados por MCKENZIE (1985). En el presente trabajo se utiliza un modelo INAR(1) con marginal Poisson, cuyas características son detalladas en la siguiente sección.

2 DESARROLLO

2.1 Descripción del Modelo INAR(1)

El modelo INAR(1) tiene la siguiente estructura:

$$x_t = \alpha \circ x_{t-1} + \varepsilon_t$$

donde “ \circ ” representa el operador de refinamiento binomial (*binomial thinning*) de forma tal que $\alpha \circ x_{t-1} = \sum_{i=1}^{x_{t-1}} y_i$, siendo las y_i variables aleatorias Bernoulli con probabilidad de éxito α , independientes entre sí y también de x . Por otra parte la operación de refinamiento es efectuada en forma independiente en cada momento t y los ε_t son i.i.d. y siguen una distribución discreta de probabilidad.

Si bien el modelo parece complejo, es bastante intuitivo. Por ejemplo: supongamos que $x_{t-1} = 3$ y $\alpha = 0,2$, aplicando el operador de refinamiento $\alpha \circ x_{t-1} = \sum_{i=1}^{x_{t-1}} y_i$, se tiene que $2 \circ 3 = \sum_{i=1}^3 y_i$, donde $y_i \sim B(1; 0,2)$, entonces se generan tres observaciones Bernoulli con probabilidad de éxito igual a 0,2 y se las suma, dicha suma tendrá un valor esperado de $0,2 \times 3 = 0,6$, sin embargo el proceso de refinamiento garantiza que obtendremos un valor entero.

Por último para el término de error se pueden usar varias distribuciones discretas, de las cuales en el presente trabajo se tendrá $\varepsilon_t \sim P(\lambda)$. Esto se conoce como un modelo INAR(1) con marginal Poisson.

Entonces se aprecia que el modelo INAR(1) permite generar datos de tipo autorregresivo con dos características: los datos son enteros y podemos utilizar errores con distribución no normal, en particular con distribución Poisson. De estas apreciaciones se deriva su utilidad para nuestra investigación, dado que permite generar datos similares a los de un proceso Poisson pero con existencia de autocorrelación.

2.2 Propiedades del Modelo INAR(1) con Marginal Poisson

En esta sub sección se enumeran algunas propiedades importantes del modelo INAR (1) con marginal Poisson. Para mayores detalles se puede consultar Al-Osh y Alzaid (1987) y Weiß (2007).

- a) Si $\varepsilon_t \sim P(\lambda)$, entonces $x_t \sim P\left(\frac{\lambda}{1-\alpha}\right)$. Esto implica en forma directa que $E(x) = Var(x) = \mu = \frac{\lambda}{1-\alpha}$.
- b) El coeficiente de correlación del modelo INAR(1) es igual al de un modelo AR(1), esto es: $\rho = \alpha$. Por lo tanto, para estimar el coeficiente de correlación, se puede utilizar el coeficiente de autocorrelación muestral de primer orden:

$$r = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t-1} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

De la propiedad b) se desprende que modificando el parámetro autorregresivo α podríamos modelar cambios en la autocorrelación. Por otra parte, usando también la propiedad a) se sigue que cambios en la autocorrelación generan cambios en la media y varianza del proceso.

2.3 Aplicación al Control Estadístico de Procesos

La gráfica “c” requiere independencia en las observaciones. Este supuesto no se cumple si la variable que se controla sigue un proceso INAR(1). Por otra parte, el supuesto de independencia se verifica en los ε_t .

Supongamos que nuestros datos siguen un proceso INAR(1). Continuando la práctica habitual construiríamos una gráfica “c” con parámetro μ , en base a los valores observados usando el estimador $\hat{\mu} = \frac{\sum_{t=1}^n x_t}{n}$, los límites de control se obtienen como²;

$$\begin{aligned} LSC &= \hat{\mu} + 3\sqrt{\hat{\mu}} \\ LC &= \hat{\mu} \\ LIC &= \hat{\mu} - 3\sqrt{\hat{\mu}} \end{aligned}$$

Debe notarse que procediendo de este modo, en forma implícita estaríamos asumiendo independencia, siendo que en realidad los datos están autocorrelacionados. Entonces, resulta de interés conocer cómo funciona el gráfico bajo estas condiciones.

Para analizar qué ocurre con el gráfico de control cuando se parte de una situación donde se cumple el supuesto de independencia y luego se pasa a una situación con autocorrelación, se puede realizar un ejercicio sencillo. Este consiste en generar cierta cantidad de observaciones en base a una distribución Poisson con parámetro λ , utilizar esos datos para construir los límites de la gráfica de control y luego generar otra serie de datos originarios de un proceso INAR(1) con marginal Poisson(λ) y parámetro autorregresivo $\alpha > 0$.

² En caso de que el límite inferior de control resulte negativo, el mismo será tomado como igual a cero.

En este punto, es importante notar que la generación de datos en base a una distribución Poisson (λ), es equivalente a la generación de datos usando un proceso INAR(1) con marginal Poisson (λ) y parámetro de autocorrelación $\alpha = 0$.

En la FIGURA N° 1, se muestra el resultado de aplicar la metodología planteada para el caso $\lambda = 5$ y $\alpha = 0,5$. Los primeros 100 datos son generados con distribución Poisson ($\lambda = 5$). En base a esa serie se construyen los límites de control y luego se incorpora una segunda serie de 100 datos, generada a partir de un proceso INAR (1) con parámetro de autocorrelación $\alpha = 0,5$ y cuyo término de error tiene distribución marginal Poisson ($\lambda = 5$).

Resulta evidente que al aumentar el parámetro de autocorrelación, manteniéndose λ constante, el proceso sale de control, siendo esto evidenciado porque varias observaciones correspondientes a la segunda serie superan el límite superior de control.

La salida de control encontrada es consecuencia directa de las propiedades del proceso INAR (1). Al aumentar el coeficiente autorregresivo, aumenta la autocorrelación, pero por otra parte también se produce un aumento en media y varianza del proceso, lo que genera la salida de control. Sin embargo, esta salida de control no constituye una falencia del gráfico, dado que el mismo detecta correctamente el aumento en la media del proceso.

Del análisis previo surge que, con el fin de aislar los efectos propios de la correlación, es deseable introducir la misma manteniendo estables la media y varianza del proceso. Con esta intención se realiza el siguiente ejercicio; se generan 100 datos con distribución Poisson ($\mu = \frac{\lambda}{1-\alpha}$). Se utilizan los mismos para fijar los límites de control y luego se generan 100 datos en base a un modelo INAR(1) con parámetro autorregresivo $\alpha > 0$ y término de error con distribución Poisson(λ). De este modo se mantienen estables la media y varianza.

En la FIGURA N° 2 se muestra la gráfica resultante de utilizar este procedimiento tomando $\alpha = 0,8$ y $\lambda = 2$. Como se puede observar, en este caso el cambio en el coeficiente autorregresivo no resulta en una salida de control del proceso. Esto indica que si aumenta la autocorrelación, manteniendo estables la media y varianza del proceso, el gráfico "c" no detecta el cambio si sólo usamos como criterio para declarar fuera de control un proceso la aparición de un dato fuera de los límites de control.

A pesar de que ambos procesos son aparentemente iguales, si calculamos los coeficientes de autocorrelación muestral encontramos que en los primeros 100 datos, el mismo es próximo a cero, mientras que los últimos 100 datos poseen un coeficiente de autocorrelación muestral próximo a 0,8. Esto indica que en este caso el coeficiente de autocorrelación muestral detecta adecuadamente la presencia de autocorrelación.

Para que un proceso se encuentre bajo control estadístico se requiere que no existan causas asignables de variabilidad, es decir que toda variabilidad se deba a la aleatoriedad. Como se afirma en MONTGOMERY (2005), aún si todos los puntos caen dentro de los límites de control, cuando estos se comportan en forma sistemática o no aleatoria existen causas asignables de variabilidad y el proceso se declara fuera de control.

Se han desarrollado reglas de sensibilización de los gráficos de control, entre ellas las tratadas en el manual de la Western Electric Company (1956), principalmente tratando de mejorar la capacidad de los gráficos para detectar cambios en la media de un proceso. Sin embargo, algunas reglas apuntan a detectar patrones no sistemáticos. En particular, resulta interesante analizar si la regla consistente en declarar fuera de control a un proceso cuando se observan ocho valores consecutivos por encima o por debajo de la línea central logra detectar la presencia de autocorrelación. El razonamiento implícito es que si los datos están autocorrelacionados, la ocurrencia de un valor a un lado de la línea central, será probablemente seguido por otro valor en la misma zona. En lo que sigue, a esa regla se la denomina "regla de rachas".

En la FIGURA N° 3, se muestran las mismas series que en la FIGURA N° 2, pero resaltando en color amarillo aquellos puntos que indican una violación de la regla de rachas. En dicho gráfico se nota que la presencia de autocorrelación produce un aumento en las rachas. En el caso de los primeros 100 datos en que no hay autocorrelación, en ningún caso se encontraron 8 valores seguidos por encima o debajo de la línea central, en cambio en los últimos 100 datos, la presencia de autocorrelación generó que se encontraran 34 casos de 8 valores seguidos por encima o debajo de la línea central.

2.4 Evaluación del Desempeño de la regla de Rachas por Simulación

Para analizar como varía la capacidad de la regla de rachas para detectar autocorrelación, se ha realizado un estudio de simulación construyendo los ARL (*average run length*) de dicha regla para varios valores de los parámetros μ y α . El ARL indica la cantidad de datos que se obtendrán, en promedio, antes de encontrar una racha de ocho valores consecutivos por encima o debajo de la línea central. En cada caso se han generado 10000 series de 1000 observaciones cada una, registrando en cuantas oportunidades se encuentra una racha y calculando que proporción de los datos representan esos casos.

Luego se tomó recíproca sobre cada una de las proporciones del paso previo, con lo que se han obtenido 10000 RL (*run length*), los cuales indican la cantidad de datos obtenidos antes de la ocurrencia de una racha en cada una de las corridas de la simulación. Finalmente, se promediaron los RL para encontrar el ARL. El mismo procedimiento fue repetido para cada configuración paramétrica. Los resultados se muestran en la TABLA N° 1.

Como es conocido, el ARL es la media de una variable aleatoria con distribución geométrica, siendo en este caso la probabilidad de éxito dada por la probabilidad de encontrar una racha de ocho valores por encima o debajo de la línea central.

Para que la regla de rachas sea adecuada, es deseable que su ARL sea pequeño cuando no existe autocorrelación, es decir cuando $\alpha = 0$. Por otra parte el ARL de la regla de rachas debe reducirse a medida que aumenta α , porque esto indicaría que la regla logra detectar cambios en la autocorrelación del proceso.

En la tabla se observa que para todos los valores de la media del proceso los ARL decaen claramente a medida que aumenta el parámetro autorregresivo. Esto indica que la regla de rachas planteada es efectiva para detectar la presencia de autocorrelación en el proceso así como los cambios en la misma.

Como crítica a la adopción de la regla de rachas, en la TABLA N° 1 se observa que su aplicación puede generar una considerable cantidad de falsas alarmas. Este problema es importante especialmente cuando el parámetro μ es chico. Por ejemplo: cuando $\mu = 2$ y $\alpha = 0$, el ARL es 50,89. Esto significa que en promedio cada aproximadamente 51 observaciones encontraremos una violación de la regla de rachas, siendo que en realidad la autocorrelación del proceso es cero. Para tratar este problema dos posibles soluciones son: utilizar una reglada rachas que requiera una cantidad mayor de puntos por encima o debajo de la línea central; o ante la aparición de una racha tomar la misma como señal de prevención y no detener el proceso.

3 CONCLUSIONES

Se ha empleado un modelo INAR(1) con media μ para generar datos autocorrelacionados enteros. Por otra parte también se generaron datos con distribución Poisson con media μ . Ambas series se usaron para observar que ocurre en la gráfica "c" cuando introducimos correlación y mantenemos inalteradas la media y varianza del proceso. Del análisis se concluye que:

- a) La presencia de correlación no ha modificado la probabilidad de encontrar observaciones fuera de los límites de control. Esto tiene como consecuencia positiva que el gráfico “c” sigue funcionando en forma similar que en el caso sin correlación para detectar cambios en la media y varianza del proceso.
- b) Si sólo se declara un proceso como fuera de control cuando encontramos puntos fuera de los límites de control, no detectaremos una salida de control debida a un cambio en la autocorrelación del proceso. Por esta razón es conveniente aplicar algún criterio extra que aborde esa problemática.
- c) Un criterio que consigue detectar los cambios en la autocorrelación es una de las clásicas reglas de sensibilización, la que consiste en observar si existe una “racha” de ocho puntos seguidos por encima o debajo de la línea central. Los ARL correspondientes a la misma caen marcadamente cuando aumenta la correlación.
- d) La cantidad de falsas alarmas debidas a la regla de rachas aumenta considerablemente cuando el parámetro μ es pequeño. En estos casos es recomendable aumentar la cantidad de valores consecutivos por encima o debajo de la línea central que utiliza la regla.

FIGURAS

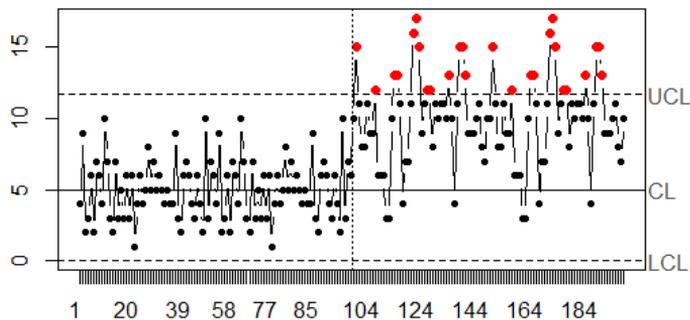


FIGURA N° 1: Comparación de Proceso Poisson($\lambda = 5$) con un proceso INAR(1) con parámetros $\lambda = 5$ y $\alpha = 0,5$.

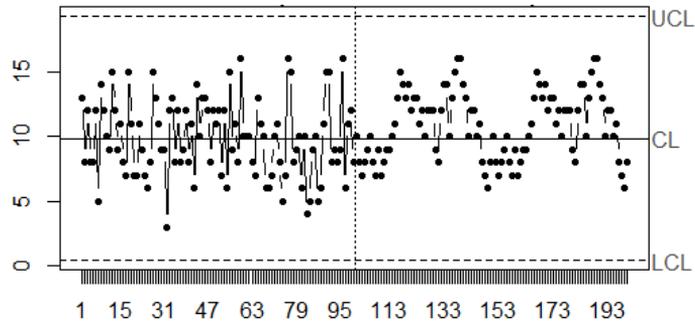


FIGURA N° 2: Comparación de Proceso Poisson($\lambda = 10$) con un proceso INAR(1) con parámetros $\lambda = 2$ y $\alpha = 0,8$.

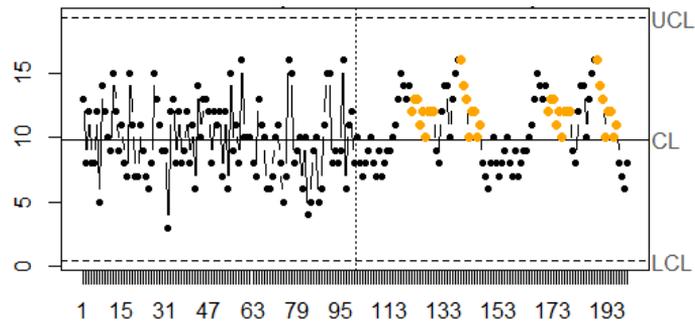


FIGURA N° 3: Comparación de Proceso Poisson($\lambda = 10$) con un proceso INAR(1) con $\lambda = 2$ y $\alpha = 0,8$ usando la regla de rachas.

TABLAS

		Valores de α			
		0	0.2	0.5	0.8
Valores de μ	2	50.89	26.16	8.80	2.80
	4	74.96	34.19	10.90	3.16
	6	84.86	38.19	11.78	3.30
	8	92.85	41.85	12.28	3.37
	10	98.54	43.88	12.56	3.41

TABLA N° 1: ARL para las distintas configuraciones paramétricas simuladas.

REFERENCIAS

- AL-OSH, M.; ALZAID, A. (1987): "FIRST-ORDER INTEGER-VALUED AUTORREGRESIVE (INAR(1)) PROCESS". *Journal of Time Series Analysis* – Vol. 8(3) – pgs. 261, 275.
- ALWAN, L. C.; ROBERTS, H. V. (1988): "TIME SERIES MODELING FOR STATISTICAL PROCESS CONTROL". *Journal of Business and Economic Statistics* – Vol. 6(1) – pgs. 87, 95.
- MCKENZIE, E. (1985): "SOME SIMPLE MODELS FOR DISCRETE VARIATE TIME SERIES". *Water Resources Bulletin* – Vol. 21(4) – pgs. 645, 650.
- MONTGOMERY, D. (2005): *Introduction to Statistical Quality Control* – Fifth Edition, John Wiley and Sons.
- WEIß, C. H. (2007): "CONTROLLING CORRELATED PROCESSES OF POISSON COUNTS". *Quality and Reliability Engineering International* – Vol. 23(6) – pgs. 741, 754.
- WESTERN ELECTRIC COMPANY (1956): *Statistical Quality Control Handbook*, Indianapolis.