



UNIVERSIDAD NACIONAL DE COLOMBIA

# Monitoreo de Procesos con Datos Censurados

**Romario Ademir Conto López**

Universidad Nacional de Colombia  
Facultad de Ciencias, Escuela de Estadística  
Medellín, Colombia  
2016

# Monitoreo de Procesos con Datos Censurados

Romario Ademir Conto López

Tesis presentada como requisito parcial para optar al título de:  
**Magíster en Ciencias-Estadística**

Directora:  
Nelfi G. González Álvarez  
Doctora en Ciencias Estadísticas

Línea de Investigación:  
Control Estadístico de Procesos  
Grupo de Investigación:  
Estadística Industrial

Universidad Nacional de Colombia  
Facultad de Ciencias, Escuela de Estadística  
Medellín, Colombia  
2016

## Dedicatoria

A mis padres, Luz Marina López y Wilfrido Conto, quienes siempre me apoyaron y me dieron la oportunidad de estudiar.

“Los Estadísticos son patriotas, luchan por la independencia.”

Anónimo.

# Agradecimientos

Agradezco a mis padres, ya que sin su apoyo jamás hubiese sido posible desarrollar esta tesis. A mi asesora, la profesora Nelfi González, quien con sus acertada guía permitió la culminación de este arduo trabajo. También agradezco a la Universidad Nacional de Colombia y en particular a la Escuela de Estadística de la sede Medellín, la cual me abrió la puertas para desarrollar mis estudios de posgrado. Y en particular agradezco al "Programa Nacional de Apoyo a Estudiantes de Posgrado para el fortalecimiento de la Investigación, Creación e Innovación de la Universidad Nacional de Colombia 2013-2015", el cual me brindó un gran apoyo económico y anímico durante el desarrollo de esta tesis.

## Resumen.

En ocasiones durante el monitoreo de procesos no es posible realizar mediciones exactas sobre las unidades observadas, por cuestiones bien sea de costos y/o tiempo, por tanto es necesario trabajar con observaciones censuradas y por ello no es apropiado aplicar cartas de control tipo Shewhart tradicionales ya que la censura disminuye la potencia para detectar cambios en los parámetros del proceso que se está monitoreando. Ante esta situación, es recomendable la implementación de cartas de control que consideren el tipo de censura y la distribución estadística de los datos, con el fin de detectar rápidamente posibles cambios en el proceso que puedan ocasionar situaciones de fuera de control estadístico.

**Palabras clave:** Censura, Monitoreo de procesos, Cartas de control, Valor Esperado Condicional.

## Abstract.

Sometimes during process monitoring is not possible take exact measurements of the units observed, therefore it should be necessary to work with censored observations. Then it is not appropriate a direct application of traditional Shewhart charts, because in the presence of censoring there is a lower power to detect changes in the process parameters being monitored. According to this situation, we consider the implementation of control charts to take in to account the censorship and a statistical distribution for the data, with the purpose of detect, quickly, possible changes in the process that can cause situations out of statistical control.

**Keywords:** Censoring, Monitoring process, Control Charts, Conditional Expected Value.

# Contenido

<b>Agradecimientos</b>	<b>IV</b>
<b>Resumen</b>	<b>v</b>
<b>1. Introducción</b>	<b>2</b>
<b>2. Marco Teórico</b>	<b>4</b>
2.1. Control de calidad . . . . .	4
2.2. Cartas de control . . . . .	4
2.3. Censura . . . . .	5
2.4. Estimación por máxima verosimilitud. . . . .	7
<b>3. Estado del Arte</b>	<b>9</b>
<b>4. Carta <math>\bar{X}</math> con Censura a Intervalo</b>	<b>16</b>
4.1. Carta de Control CEV- $\bar{X}$ con censura a intervalo . . . . .	17
4.2. Carta de control $\bar{X} - b$ con censura a intervalo. . . . .	22
4.3. Carta de control $\bar{X} - U$ con censura a intervalo . . . . .	24
4.4. Potencia para las cartas $\bar{X}$ propuestas . . . . .	26
4.5. Potencia comparativa de las cartas $\bar{X}$ propuestas Vs. la carta tipo Shewhart tradicional . . . . .	31
4.6. Caso simulado . . . . .	35
<b>5. Trabajo futuro, conclusiones y recomendaciones</b>	<b>39</b>
5.1. Estimación de parámetros . . . . .	39
5.1.1. Estimación por máxima verosimilitud . . . . .	39
5.1.2. Estimación por métodos de optimización . . . . .	44
5.2. Conclusiones . . . . .	47
5.3. Trabajo futuro y recomendaciones . . . . .	48
<b>A. Optimización usando el paquete “DEoptim” en R</b>	<b>49</b>
<b>B. Optimización usando función “optim” en R</b>	<b>51</b>

# Lista de Símbolos

$P_c$	Proporción de censura de acuerdo al nivel de censura $c$
$P_{a,b}$	Proporción de censura de acuerdo al intervalo de censura $a,b$
$Q(\cdot)$	Función de supervivencia normal estándar
$F(\cdot)$	Función de distribución normal $\mu, \sigma^2$
$\bar{F}(\cdot)$	Función de supervivencia normal $\mu, \sigma^2$
$\phi(\cdot)$	Función de densidad normal estándar
$w_c$	Valor esperado condicional bajo normalidad y censura a derecha tipo I
$w_{a,b}$	Valor esperado condicional bajo normalidad y censura a Intervalo tipo I

# Lista de Figuras

2-1. Ejemplo carta $\bar{X}$ Shewhart en control estadístico. Elaboración propia. . . . .	5
4-1. UCLs y LCLs Carta CEV- $\bar{X}$ . . . . .	21
4-2. UCLs y LCLs Carta $\bar{X} - b$ . . . . .	23
4-3. UCLs y LCLs Carta $\bar{X}$ imputando con una U(a,b). . . . .	25
4-4. Potencias comparativas entre cartas propuestas con $p = 0.5$ . (a) $n = 3$ ; (b) $n = 5$ ; (c) $n = 10$ ; (d) $n = 20$ . . . . .	28
4-5. Potencias comparativas entre cartas propuestas con $p = 0.9$ . (a) $n = 3$ ; (b) $n = 5$ ; (c) $n = 10$ ; (d) $n = 20$ . . . . .	29
4-6. Potencias comparativas entre cartas propuestas con $p = 0.95$ . (a) $n = 3$ ; (b) $n = 5$ ; (c) $n = 10$ ; (d) $n = 20$ . . . . .	30
4-7. Potencia comparativa entre la carta Shewhart y las propuestas con $p = 0.5$ . (a) $n = 3$ ; (b) $n = 5$ ; (c) $n = 10$ ; (d) $n = 20$ . . . . .	32
4-8. Potencia comparativa entre la carta Shewhart y las propuestas con $p = 0.9$ . (a) $n = 3$ ; (b) $n = 5$ ; (c) $n = 10$ ; (d) $n = 20$ . . . . .	33
4-9. Potencia comparativa entre la carta Shewhart y las propuestas con $p = 0.95$ . (a) $n = 3$ ; (b) $n = 5$ ; (c) $n = 10$ ; (d) $n = 20$ . . . . .	34
4-10. Cartas propuestas aplicadas a caso simulado con cambio en escalón para la media. (a) Carta CEV- $\bar{X}$ ; (b) Carta $\bar{X} - b$ ; (c) Carta $\bar{X} - U$ ; (d) Carta $\bar{X}$ Shewhart . . .	37
4-11. Cartas propuestas aplicadas a caso simulado con cambio gradual en la media. (a) Carta CEV- $\bar{X}$ ; (b) Carta $\bar{X} - b$ ; (c) Carta $\bar{X} - U$ ; (d) Carta $\bar{X}$ Shewhart . . . . .	38



# Lista de Tablas

4-1. Límites de control, sin aplicar suavizamiento spline, por carta para proporciones de censura $p = 0.5, 0.9$ y $0.95$ y tamaños de muestra $n = 3, 10$ y $20$ . . . . .	27
4-2. Límites de control simulados para la carta Shewhart ignorando censura. . . . .	31
5-1. Parámetros estimados mediante algoritmos evolutivos y sus sesgos, con censura $p = 0.5, 0.9$ y $0.95$ y tamaños de muestra $n = 3, 5, 10$ y $20$ . . . . .	46

# 1 Introducción

En ciertos casos el monitoreo de procesos trabaja con datos censurados, y bajo esa situación no se conoce el valor exacto de la característica de calidad observada, básicamente debido a que durante el estudio existen limitaciones de tiempo, de costos, o en la precisión del instrumento de medición, y por tanto no es posible conocer con exactitud el valor de la variable o característica monitoreada. En este escenario se debería tener en cuenta cuál técnica de control usar para el manejo de este tipo de observaciones. Para ello deben ser definidos la clase y el nivel de censura, los cuales en ciertas circunstancias son fijados por parte del investigador en el problema específico, pero en otros casos el nivel o valor de censura puede ser aleatorio.

Steiner y Mackay (2000) plantean un caso en el cual aparece censura ocasionada por limitaciones en un ensayo de resistencia de materiales, en tal escenario no resulta eficiente trabajar con los cartas de control tipo Shewhart para la media y la desviación estándar, ya que su potencia para detectar cambios en el parámetro de interés (la media de resistencia), disminuye a medida que aumenta el nivel de censura; debido a esto es más adecuada la aplicación de cartas que introduzcan la información de la censura en la estimación de los parámetros distribucionales de la variable monitoreada, permitiendo extraer toda la información disponible en la muestra y se proceda a la imputación de los datos censurados.

En un segundo caso, también en Steiner y Mackay (2000), son usadas barras mecanizadas con un diámetro determinado por los límites de especificaciones en el proceso para controlar el tamaño de las arandelas en la producción de toma corrientes. Si la barra con mayor diámetro entra en el orificio, entonces se ha excedido el límite de especificación superior, y si la más pequeña no entra en el orificio, el diámetro de ésta es menor que la especificación inferior. Para las unidades que no están dentro de los límites se mide el diámetro, en tanto que las unidades dentro de límites quedan censuradas. Este tipo de censura es conocida como censura tipo I a intervalo.

Este trabajo de tesis tiene como objetivo: proponer una metodología para la construcción de una carta de control tipo Shewhart que permita monitorear la media del proceso cuando se tiene censura tipo I a intervalo y que la variable  $X$  a monitorear sigue una distribución normal con media  $\mu$  y varianza  $\sigma^2$  ( $X \sim N(\mu, \sigma^2)$ ); y también evaluar el desempeño de la carta propuesta en la detección de desviaciones del proceso rápidamente en Fase II en

comparación con la carta tipo Shewhart usual. Para ello, se sigue la metodología propuesta por Steiner y Mackay (2000), pero para el caso de censura a intervalo, donde cada valor censurado en un intervalo  $[a, b]$  es reemplazado por el valor esperado condicional (CEV), denotado por  $w_{a,b}$ , tal que  $w_{a,b} = E[X|a < X < b]$ . Adicionalmente, serán usados otros dos métodos de imputación: el límite superior de censura y el valor aleatorio de una uniforme en el intervalo  $[a, b]$  de censura. Por tanto, se proponen tres cartas de control bajo este escenario de censura. Se pretende medir el efecto que tiene la censura a intervalo sobre la capacidad de las tres cartas para detectar señales cuando la variable medida tiene distribución normal. Cabe aclarar que sólo se considera el problema del monitoreo en fase II de control, donde se asumen parámetros conocidos para la distribución estadística subyacente de la característica de calidad de interés, debido a lo cual se elige estudiar la potencia de las cartas propuestas, pues bajo parámetros conocidos dicha potencia es el inverso del ARL.

El desarrollo de esta Tesis se inicia en el Capítulo 2 con el marco teórico, donde se resaltan conceptos importantes a considerar, como la definición de cartas de control y censura. El Capítulo 3 contiene el estado arte sobre el tratamiento de la censura que diversos autores han dado durante el monitoreo de procesos. En el Capítulo 4 se desarrollan las tres cartas propuestas, las cuales difieren en el método de imputación, para el caso de censura a intervalo tipo I en fase II de control, como también son comparadas las cartas  $\bar{X}$  propuestas y la tipo Shewhart tradicional, además al final del capítulo se presenta un caso simulado con el fin de estudiar el funcionamiento de las cartas desarrolladas y la tipo Shewhart para datos fuera de control. El Capítulo 5 son descritos como trabajo futuro tres métodos de estimación de parámetros cuando estos no son conocidos, es decir, cuando se está en fase I; también se enuncian las conclusiones y recomendaciones de esta tesis.

## 2 Marco Teórico

En este capítulo se definen conceptos referentes al tema de control estadístico de calidad, describiendo el funcionamiento de las cartas de control como herramientas gráficas de interés para el monitoreo de procesos. También se define en qué consiste la censura, sus tipos e implicaciones sobre la eficiencia del esquema de monitoreo y finalmente se expone que al estimar por máxima verosimilitud parámetros desconocidos durante el proceso, se debe tener en cuenta que la función de verosimilitud cambia de acuerdo al tipo de censura presente.

### 2.1. Control de calidad

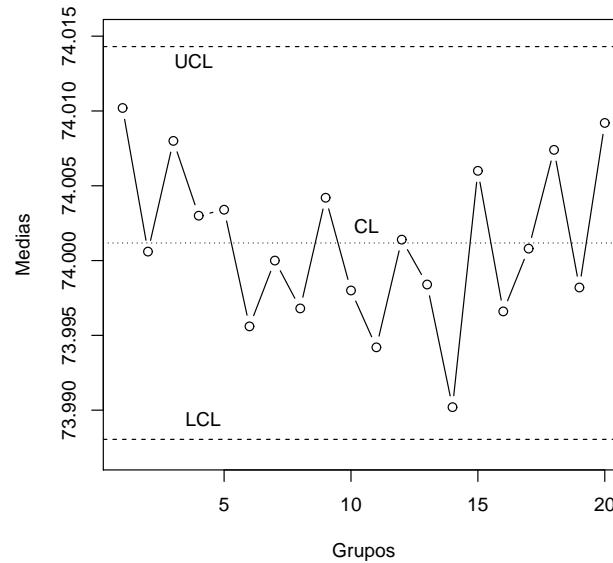
El concepto de calidad se relaciona generalmente con el cumplimiento de las especificaciones de un proceso, y por tanto, mientras se esté dentro de unos límites establecidos existirá calidad (Woodall, 2000). Para el ingeniero y estadístico japonés Genichi Taguchi (1986), la calidad no es sólo cumplir con estándares, sino que ésta es realmente lograda con la reducción de la variabilidad del proceso.

El monitoreo o control estadístico de procesos (SPC) consiste en el uso de métodos estadísticos para el estudio y mejora en la eficiencia de los procesos en el tiempo (Woodall, 2000). Se considera que un proceso está bajo control estadístico cuando la distribución de probabilidad de una característica de calidad de interés es constante en el tiempo, y si dichas condiciones no se cumplen, el proceso estaría fuera de control (Woodall, 2000). Por tanto, el SPC es un conjunto de técnicas mediante las cuales se vigila que el comportamiento de una variable de interés esté dentro de un rango que permite clasificar al producto como conforme a las características de calidad y tomar correctivos si no lo está. Siendo así, los objetivos del control de proceso son (Woodall, 2000):

- Minimizar la producción defectuosa.
- Comparar la producción con las especificaciones.
- Mantener la mejora continua del proceso.

### 2.2. Cartas de control

Representan una herramienta gráfica donde una variable o característica de calidad es medida en el tiempo, y se estudia si su valor observado en cada muestra está dentro de unos límites



**Figura 2-1:** Ejemplo carta  $\bar{X}$  Shewhart en control estadístico. Elaboración propia.

de control (Woodall, 2000). Dichos gráficos o cartas de control permiten distinguir entre dos fuentes de variación en el valor de una característica de calidad estudiada, una debida a causas asignables o especiales y otra a causas comunes (Woodall, 2000). Las causas asignables hacen referencia a eventos inusuales en el proceso, como por ejemplo errores de medición, y las causas comunes son propias del proceso, aleatorias y más difíciles de detectar.

Los gráficos o cartas de control conocidos como tipo Shewhart se caracterizan por tener una línea central, la cual representa la media de una característica de calidad o de un estadístico que está siendo monitoreado; también poseen unos límite inferior y un límite superior de control; los cuales delimitan la región de aceptación para la hipótesis durante el monitoreo donde el proceso está bajo control o dentro de los estándares deseados (Woodall, 2000). Uno de los gráficos tipo Shewhart más usados es la carta  $\bar{X}$ , la cual es construida con el fin de monitorear la media de la característica de calidad a controlar. En la Figura 2-1 se ilustra una carta  $\bar{X}$ , en la cal se puede apreciar que el proceso está bajo control, pues los puntos que representan las medias de subgrupo racional están dentro de límites .

## 2.3. Censura

El término censura es comúnmente usado en la teoría de confiabilidad y análisis de datos, en relación a la imposibilidad de observar el tiempo de vida exacto de una unidad o producto en funcionamiento Crowder et al. (1994). Por tanto, es posible que de acuerdo al esquema de monitoreo implementado se presente censura en las observaciones. Cabe notar que existen

diferentes tipos de censura, que Crowder et al. (1994) define como:

**Censura Tipo I:** En este caso se toma una muestra de  $n$  artículos que son sometidos a estudio durante un período de tiempo  $t$  fijo para todos, en el cual se observa el tiempo de vida de quienes fallan antes de  $t$ , donde todas las observaciones que no fallaron quedan censuradas.

**Censura Tipo II:** Son tomados  $n$  artículos desde un tiempo  $t = 0$  y se observa el tiempo de funcionamiento de estos hasta que se presentan  $\gamma$  fallos, cantidad previamente especificada. En este caso la cantidad de fallos será fija y el tiempo de observación total aleatorio, de modo que el tiempo de vida de las unidades que fallan después de la cantidad  $\gamma$  queda censurado aleatoriamente.

**Censura Tipo III:** En este caso, son definidos tanto el tiempo  $t$  límite hasta el cual son tomadas las mediciones, como la cantidad de fallos  $\gamma$  que serán observados, donde al cumplirse cualquiera de las dos condiciones termina el experimento.

La censura no sólo aparece en términos de tiempos de vida, sino que también se presenta en otros contextos, como en estudios relacionados con la resistencia de materiales a ciertas cargas y demás características físicas de productos que son de interés para la calidad. Dicha censura puede ser clasificada en: a derecha, a izquierda o a intervalo, de acuerdo a la dirección en que se encuentre la censura en las observaciones a la hora de tomar los datos y que depende del nivel de censura fijado en el estudio.

**Censura a Derecha:** Se caracteriza por que no es posible observar los valores que toma la variable después de sobrepasar cierto nivel, conocido como nivel de censura.

**Censura a Izquierda:** Se presenta cuando al iniciarse el estudio se evidencia que la unidad estudiada ya ha fallado y por tanto se desconoce en qué valor de la variable medida ocurrió dicha falla.

**Censura a Intervalo:** Se evidencia cuando en la unidad observada la variable estudiada toma valores dentro de un intervalo de inspección fijado.

En cuanto a la imputación de datos censurados, en el caso de censura a intervalo, comúnmente en la práctica cada valor censurado es reemplazado por el límite superior de censura, y así, se asume por conocido el valor de la variable de interés aunque no se conoce realmente; también se proponen otros métodos de imputación para la censura, como el descrito por Steiner y Mackay (2000), mediante el llamado Valor Esperado Condicional (CEV) y el expuesto por Hsu, et al. (2006), usando un valor aleatorio de una uniforme que tiene como parámetros los límites inferior y superior de censura.

## 2.4. Estimación por máxima verosimilitud.

La estimación por máxima verosimilitud es uno de los métodos estadísticos más usados en la estimación de parámetros, y consiste en maximizar la función de densidad conjunta de una muestra aleatoria  $X_1, X_2, \dots, X_n$ , con respecto a los parámetros de la función de densidad para la variable aleatoria cuando se tiene una realización de la muestra, asumiendo una distribución estadística para la población.

Suponga que se tiene una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una población, donde  $X_i$  son los valores observados de la variable aleatoria  $X$ , y considerando inicialmente que no hay censura. Sea  $f(x; \theta)$  la función de densidad conocida, con parámetros  $\theta = \theta_1, \dots, \theta_m$  desconocidos, los cuales son estimados con los datos observados, entonces la función de verosimilitud será:

$$L(\theta) = C \prod_{i=1}^n f(x_i, \theta) \quad (2-1)$$

Así son hallados los valores de  $\theta$  que maximizan  $L(\theta)$ . Usualmente la constante  $C$  no depende de  $\theta$ , puede ser simplificada con  $C = 1$  para la estimación de  $\theta$  (Meeker y Escobar, 1998).

Las contribuciones a la verosimilitud de la censura a intervalo, a derecha y a izquierda son simplemente la probabilidad que la variable esté en el respectivo intervalo de estudio (Meeker y Escobar, 1998). Si el valor de la variable está entre  $x_{i-1}$  y  $x_i$ , la probabilidad de este evento es:

$$L_i(P) = \int_{x_{i-1}}^{x_i} f(x) dt = F(x_i) - F(x_{i-1}) \quad (2-2)$$

Para el caso de censura a izquierda se sabe que el valor de la variable es menor que un valor dado a partir del cual se comienza a medir exactamente la característica de calidad de una unidad, entonces, la contribución a la probabilidad y verosimilitud de cada observación censurada es:

$$L_i(P) = \int_0^{x_i} f(x) dt = F(x_i) - F(0) = F(x_i) \quad (2-3)$$

Considerando censura a derecha, la censura se presenta cuando el valor de la característica de calidad de un elemento es mayor que un valor definido, es decir que el valor de la variable está en el intervalo  $(x_i, \infty)$ , por tanto la contribución a la verosimilitud de cada observación con este tipo de censura es:

$$L_i(P) = \int_{x_i}^{\infty} f(x) dt = F(\infty) - F(x_i) = 1 - F(x_i) \quad (2-4)$$

Finalmente la función de verosimilitud total o de probabilidad conjunta para la muestra, con  $n$  observaciones independientes está dada por (Meeker y Escobar, 1998),

$$L(\theta) = C \prod_{i=1}^n [f(x_i)]^{m_i} [F(x_i)]^{l_i} [F(x_i) - F(x_{i-1})]^{d_i} [1 - F(x_i)]^{r_i} \quad (2-5)$$

Donde  $n = \sum_{i=1}^n (m_i + l_i + d_i + r_i)$ ,  $m_i$  es el número de observaciones exactas,  $l_i$  es el número de observaciones con censura a izquierda,  $d_i$  la cantidad de observaciones con censura a intervalo y  $r_i$  las de censura a derecha;  $C$  es una constante que depende de la muestra mas no del vector de parámetros  $\theta$ , y puede ser igualada a 1.



## 3 Estado del Arte

En este capítulo se lleva a cabo una breve revisión en la literaria sobre algunos desarrollos y hallazgos presentados por diferentes autores sobre el tema de la censura en el monitoreo de procesos. Para ello, se describen algunos artículos relevantes, para luego centrarse en Steiner y Mackay (2000) y Vargas y Montaña (2005), ya que ambos exponen procedimientos desarrollados con objetivos similares a los de esta tesis pero en el caso de censura a derecha. El tema de la censura en el monitoreo de procesos ha sido estudiado previamente. Por ejemplo, Feltz y Sturm (1994) describen una técnica bayesiana empírica en la construcción de cartas para monitorear características de un proceso cuando se presenta censura a derecha, a izquierda o ambas. También la propuesta es aplicable a datos censurados por intervalo, donde las observaciones quedan censuradas dentro de un intervalo determinado, obteniéndose mediciones exactas sólo por encima del límite superior y por debajo del límite inferior. La propuesta permite monitorear un proceso en tiempo real con el objetivo de identificar cambios en los parámetros del mismo a medida que ocurren. En este trabajo se muestra que aunque no todas las unidades pueden ser medidas exactamente, con la información resultante el proceso puede ser monitoreado y es posible identificar causas de variación. Por tanto, la metodología bayesiana implementada por Feltz y Sturm (1994) permite calcular la probabilidad de ocurrencia de una observación por fuera del intervalo en algún tiempo dado, y considerando las observaciones que no fueron censuradas se pueden realizar estimaciones para la media del proceso y aproximar la distribución inherente del proceso estudiado.

Steiner y Mackay (2001) implementan una carta basada en el Valor Esperado Condicional (CEV) de medias móviles ponderadas exponencialmente (EWMA) tomando en consideración que la censura resulta de riesgos competitivos, donde hay dos o más modos de fallo. Los autores plantean un ejemplo, donde un adhesivo se usa para unir un tejido de vinilo con la espuma para el recubrimiento interior de un automóvil. La fuerza de la unión adhesiva es una característica clave y es de interés para controlar la media de la resistencia de la unión. Sin embargo, durante las pruebas, la resistencia a la ruptura no es siempre observada, ya que el refuerzo de la espuma puede dañarse primer. Este tipo de censura ocasiona que los procedimientos de control cambien de acuerdo a la variación del denominado riesgo competitivo. En estos casos la distribución de la media muestral de los pesos CEV depende de la media y la desviación estándar del riesgo competitivo. Steiner y Mackay (2001) también implementan dos cartas EWMA bajo la distribución normal, donde en una es controlada la fuerza de resistencia media y en otra se controla la media del riesgo competitivo.

Zhang y Chen (2004), desarrollan límites inferior y superior apropiados para la carta EWMA con censura para detectar incrementos y decrecimientos en la media del proceso considerando que los datos siguen una distribución Weibull. En este caso, consideran un nivel de censura fijo tipo I a derecha, construyen ambos límites de control y luego comparan la potencia de detectar cambios persistentes en la media del proceso con la carta  $\bar{X}$  tipo Shewhart mediante la comparación de la longitud media de corridas (ARL) de ambos. Ellos concluyen en su estudio que el límite inferior construido bajo la distribución Weibull es mejor que el propuesto por Steiner y Mackay (2001), bajo distribución normal, en su capacidad de detectar decrecimientos en la media del proceso rápidamente, de igual forma funciona el límite superior considerado. La comparación pudo ser realizada con un nivel de censura bajo (del 50% o menos), ya que la carta  $\bar{X}$  tipo Shewhart no es buena para detectar desviaciones con alta censura (mayor al 90%).

Brill y Bzik (2009) realizan una propuesta para la construcción de una carta de control que permita controlar el error tipo I para datos con distribución asimétrica y considerando censura a izquierda. El desarrollo de esta carta es motivada por necesidades encontradas en la industria relacionadas con la satisfacción del cliente. Un comprador busca que los productos que consume cumplan con especificaciones nominales definidas durante el proceso productivo, y adicionalmente posean ciertas características esperadas por él; por tanto, se requieren límites de control adicionales que conlleven al cumplimiento de especificaciones más restrictivas dadas por el cliente, las cuales varían con el tiempo. Bajo este escenario se hace necesario que los límites de control sean reajustados cada cierto intervalo regular de tiempo. De modo que, si un producto es elaborado cuando el proceso está fuera de control, entonces tampoco cumplirá con los requerimientos del cliente. Brill y Bzik (2009) exponen como una necesidad económica, controlar la tasa de error tipo I o de falsa alarma para estos límites de control cambiantes. Es decir, que a medida que los límites de control sean ajustados de acuerdo a las necesidades del cliente, se conserva la tasa de error tipo I de la carta. Los límites de control son calculados teniendo en cuenta la media, la desviación estándar y el coeficiente de asimetría muestral.

Pascal y Li (2011) consideran el problema de la construcción de cartas de control que permitan monitorear el comportamiento del parámetro de forma ( $\beta$ ) de la Weibull, basadas en el rango muestral y considerando censura tipo II a derecha en los datos. Ellos exponen que el parámetro de forma juega un papel importante en las aplicaciones de la distribución Weibull en confiabilidad y análisis de supervivencia, ya que el riesgo de fallo depende del comportamiento de este parámetro. Así mismo, tratan el caso de censura tipo II a derecha, presente cuando se llevan a cabo  $n$  pruebas simultáneamente y se mide el tiempo de vida sobre las unidades de interés, finalizando la prueba cuando ocurren  $c$  fracasos. Los límites de control para la propuesta son determinados de acuerdo al número de fallas exactas  $c$ , el ARL bajo control deseado y considerando que el parámetro de forma es estable. Pascal y Li

(2011) exponen que se puede construir la carta con un sólo límite o con ambos dependiendo de si en el proceso se desea estudiar sólo el crecimiento o el decrecimiento del parámetro de forma y/o ambas situaciones.

Con relación a las cartas  $\bar{X}$ , Steiner y Mackay (2000) plantean el problema sobre la fabricación de un material utilizado en el acabado interior de un automóvil, el cual va pegado a un refuerzo de espuma aislante, y se desea medir la resistencia a la unión que tiene el material, por lo cual se corta una capa rectangular del mismo y se mide la fuerza requerida para romper la unión sin dañar la espuma, con una determinada fuerza máxima de aplicación. Si en los ensayos al aplicar la fuerza máxima no se ha roto la adhesión, para las unidades sobrevivientes su variable resistencia queda censurada en este valor; por tanto el tipo de censura será a derecha tipo I, debido a que para valores mayores a la fuerza máxima existe censura, y es de tipo I porque se ha fijado un nivel de censura en el estudio igual a la fuerza de adhesión máxima. Por tanto construyen una carta  $\bar{X}$  y  $S$  basados en el valor esperado condicional (CEV) de la variable monitoreada con censura tipo I a derecha asumiendo una distribución normal. En estas cartas y debido a que la variable medida es la resistencia máxima a la tracción, interesará estudiar si la media del proceso disminuye o la variabilidad aumenta. Por esto dichas cartas tienen un sólo límite, donde al monitorear la media se tendrá un límite inferior y al estudiar la varianza un límite superior.

Para construir las cartas, Steiner y Mackay (2000) exponen que bajo el supuesto que el proceso está bajo control, la variable fuerza  $X \sim N(\mu, \sigma^2)$  y se define un nivel de censura  $c$  a derecha, por tanto la proporción de censura ( $P_c$ ), está dada por:

$$P_c = 1 - F(c) = Q\left(\frac{c - \mu}{\sigma}\right), \quad (3-1)$$

donde  $Q(\cdot)$  corresponde a la función de supervivencia normal estándar. La idea es reemplazar cada observación censurada por el valor esperado condicional (CEV), el cual puede ser calculado considerando que:

$$F^*(x) = P(X \leq x | X > c) = \frac{P[(X \leq x) \cap (X > c)]}{P(X > c)},$$

donde,

$$P_c = P(X > c) = \bar{F}(c).$$

Hay dos casos:

$$\text{Si } x < c, \text{ entonces } P[(X \leq x) \cap (X > c)] = 0,$$

$$\text{Si } x > c, \text{ entonces } P[(X \leq x) \cap (X > c)] = P(c < X \leq x) = F(x) - F(c).$$

luego,

$$F^*(x) = \begin{cases} 0 & x \leq c \\ \frac{F(x) - F(c)}{\bar{F}(c)} & x > c, \end{cases} \quad (3-2)$$

$$f^*(x) = \frac{dF^*(x)}{dx} = \begin{cases} 0 & x \leq c \\ \frac{f(x)}{\bar{F}(c)} & x > c. \end{cases} \quad (3-3)$$

Conociendo  $f^*(x)$  se puede calcular:

$$E[X|X > c] = \int_c^\infty x f^*(x) dx = \int_c^\infty x \frac{f(x)}{\bar{F}(c)} dt = \frac{1}{\bar{F}(c)} \int_c^\infty x f(x) dx.$$

Si  $X \sim N(\mu, \sigma^2)$ , entonces  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ ,  $E[X] = \mu$ ,  $\text{Var}[X] = \sigma^2$ , y así:

$$E[X|X > c] = \frac{1}{\bar{F}(c)} \int_c^\infty x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx.$$

Sustituyendo  $u = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma u + \mu$  y  $du = \frac{dx}{\sigma}$ ,

$$E[X|X > c] = \frac{1}{Q(Z_c)} \left[ \mu \int_{Z_c}^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du + \sigma \int_{Z_c}^\infty u \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du \right],$$

donde  $Q(Z_c) = \bar{F}(c) = 1 - F(c)$ , entonces:

$$E[X|X > c] = \frac{1}{Q(Z_c)} \left[ \mu Q(Z_c) + \sigma \int_{\frac{Z_c}{2}}^\infty \frac{1}{\sqrt{2\pi}} \exp\{-v\} dv \right],$$

$$E[X|X > c] = \mu + \sigma \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{Z_c^2}{2}\right\}}{Q(Z_c)},$$

$$E[X|X > c] = \mu + \sigma \left( \frac{\phi(Z_c)}{Q(Z_c)} \right) = w_c, \quad (3-4)$$

Por tanto, la variable  $w$  queda definida como:

$$w = \begin{cases} x & \text{si } x \leq c \quad (\text{No censurado}), \\ w_c & \text{si } x > c \quad (\text{Censurado}). \end{cases} \quad (3-5)$$

Para la estimación de los parámetros  $\mu, \sigma$ , se tiene que para una muestra de tamaño  $n$ , la log-verosimilitud para datos censurados a derecha es (Steiner y Mackay, 2000):

$$\log L(\theta) = -r \log \sigma + (n - r) \log \left[ Q \left( \frac{c - \mu}{\sigma} \right) \right] + \sum_{i \in D} \log \left[ \phi \left( \frac{x_i - \mu}{\sigma} \right) \right], \quad (3-6)$$

donde  $\phi(\cdot)$  es la función de densidad acumulada normal estándar,  $r$  es el número de observaciones no censuradas y  $D$  es el conjunto de observaciones que son medidas exactamente. Ahora bien, para determinar el límite de control superior de la carta  $\bar{X}$  se debe tener en cuenta que para las cartas CEV propuestas, éste depende del tamaño de la muestra y de la proporción de censura ( $P_c$ ); para ello, fueron generados datos provenientes de una  $N(0, 1)$  para diferentes tamaños de muestra (3, 5, 10, 20), un nivel de censura dada y considerando una probabilidad de falsa alarma teórica de 0.0027, por tanto el límite de control superior está dado por la siguiente expresión:

$$\text{Límite inferior Carta CEV } \bar{X} = LCL_{\bar{Z}} * \sigma + \mu, \quad (3-7)$$

donde  $LCL_{\bar{Z}}$  es el límite de control calculado por simulación con muestras con censura a derecha tipo I, provenientes de una  $N(0, 1)$ , por tanto el límite inferior de control para datos  $N(\mu, \sigma^2)$  es obtenido al multiplicar el límite simulado por la desviación estándar muestral  $\sigma$  y luego sumarle la media del proceso  $\mu$ . Un procedimiento similar de simulación es realizado para hallar el límite superior de la carta  $S$  con diferentes tamaños de muestra (3, 5, 10, 20):

$$\text{Límite superior Carta CEV } S = UCL_S * \sigma, \quad (3-8)$$

donde  $UCL_S$  es el límite superior de control hallado con datos de una  $N(0, 1)$ , y por tanto el límite de control para  $X \sim N(\mu, \sigma^2)$  se obtiene multiplicando el límite simulado por la desviación estándar del proceso  $\sigma$ . Luego de calcular los límites y construir la carta propuesta, Steiner y Mackay (2000) estudian la eficiencia de las cartas  $\bar{X}$  y  $S$  tipo Shewhart tradicional cuando existe censura en comparación con las cartas CEV mediante la construcción de curvas de potencia, las cuales resultan de un gráfico donde en el eje  $y$  se ubica la probabilidad que tiene la carta en detectar decaimientos en la media del proceso, contra diferentes proporciones de censura. Luego se construye los gráficos de potencia para las cartas tipo Shewhart tradicional en conjunto con los de las cartas que toman en consideración la censura, permitiendo concluir que cuando existe censuramiento las cartas basadas en CEV tienen mayor probabilidad de detectar cambios en la media y la desviación estándar del proceso, al menos para tamaños de muestra  $n$  iguales a 3 y 5. Los autores también estudiaron el efecto que tiene la proporción de censura en la capacidad de detectar cambios en los parámetros de interés del proceso, ya que cuando existe mayor proporción de censura cada vez hay menos información en la muestra acerca de la variabilidad del proceso, siendo así Steiner y Mackay

(2000) encontraron que a medida que aumenta la proporción de censura la probabilidad de detectar cambios en las variables de interés para las cartas CEV es menor, pero mayor que cuando se ignora la censura.

Vargas y Montaña (2005) adaptan una metodología basada en el valor esperado condicional (CEV) de las observaciones censuradas para la carta  $\bar{X}$  bajo censura tipo I a derecha, suponiendo que el proceso se encuentra bajo control y que la variable estudiada tiene una distribución Weibull. Si se tiene que la variable aleatoria  $X$  que representa una característica de calidad bajo la Weibull entonces:

$$F(x, \alpha, \beta) = 1 - \exp \left[ - (x/\alpha)^\beta \right] \quad x > 0, \quad (3-9)$$

con  $\alpha > 0$  parámetro de escala y  $\beta > 0$  de forma. Cuya función de densidad y de supervivencia son respectivamente:

$$f(x, \alpha, \beta) = \left( \frac{\beta}{\alpha} \right) \left( \frac{x}{\alpha} \right)^{\beta-1} \exp \left\{ - (x/\alpha)^\beta \right\} \quad x > 0, \quad (3-10)$$

$$S(x, \alpha, \beta) = 1 - \left\{ 1 - \exp \left[ - (x/\alpha)^\beta \right] \right\} = \exp \left[ - (x/\alpha)^\beta \right] \quad x > 0. \quad (3-11)$$

La forma estandarizada de la distribución Weibull definida como  $X^*(x/\alpha)$  tiene función de densidad  $f(x, 1, \beta)$ , por tanto sólo depende del parámetro de forma, lo que la hace más flexible y muy usada para modelar tiempos de vida (Meeker y Escobar, 1998). Además la probabilidad de censura bajo una distribución Weibull está dada por la función de sobrevivencia evaluada en el nivel de censuramiento  $c$ :

$$P_c = \exp \left[ - (c/\alpha)^\beta \right]. \quad (3-12)$$

Una característica que posee la distribución Weibull es su relación con la distribución de mínimo valor extremo a través de una transformación logarítmica, donde si  $X$  tiene una distribución Weibull con parámetros  $\alpha, \beta$ , entonces  $Y = \log(X)$  tiene una distribución de mínimo valor extremo con  $\mu = \log(\alpha)$  y  $\sigma = 1/\beta$ ; luego, si  $\mu = 0$  y  $\sigma = 1$ , se obtiene una distribución de valor extremo estandarizada. Siendo así  $(Y - \mu)/\sigma$  será de valor extremo estándar.

Tomando en consideración dicha estandarización en términos de los parámetros correspondientes de la distribución Weibull se tiene que:

$$V = \frac{Y - \mu}{\sigma} = \frac{\log(X) - \log(\alpha)}{1/\beta} = \beta \log(X/\alpha). \quad (3-13)$$

Por tanto, el valor esperado condicional (CEV) de las observaciones censuradas derivado de la distribución estándar de valor extremo está dado por:

$$CEV = E[V|V \geq \beta_0 \log(c/\alpha_0)]. \quad (3-14)$$

De acuerdo a lo anterior, los pesos  $CEV(W)$  correspondientes en escala Weibull son obtenidos así:

$$CEV(W) = \begin{cases} x & x \leq c, \\ \alpha \exp(CEV/\beta) & x > c. \end{cases} \quad (3-15)$$

El límite inferior de control bajo la distribución de valor extremo estándar para la variable  $V$  requerido en la construcción de la carta  $CEV \bar{X}$ -Weibull es hallado mediante simulación para tamaños de subgrupo  $n = 3, 5$  tomando como proporción de falsa alarma 0.0027 y sumiendo  $\alpha_0 = \beta_0 = 1$ . Luego, el límite inferior apropiado para  $X$  en escala Weibull para cualquier valor de  $\alpha$  y  $\beta$  es obtenido mediante la relación existente entre  $V$  y  $X$  expuesta en la ecuación (3-13). Por tanto se tiene que dicho límite de control para  $X$  es igual a  $\alpha \exp(lcl/\beta)$ , donde el  $lcl$  es el límite inferior obtenido mediante simulación y seleccionado dependiendo del tamaño de muestra usado en el proceso.

Vargas y Montaña (2005) también comparan la eficiencia de la carta propuesta  $CEV$ -Weibull con las cartas tradicionales  $\bar{X}$  tipo Shewhart bajo normalidad y  $CEV$ -Normal mediante la construcción de gráficos de potencia, en los cuales se grafica la probabilidad de detectar decrecimientos en la media del proceso cuando se presenta un cambio en unidades de desviación estándar, y son evaluadas desviaciones moderadas y/o grandes en la media del proceso. De acuerdo a lo anterior, se obtiene que la propuesta bajo una distribución Weibull es aproximadamente igual de eficiente que cuando se considera una distribución normal para los datos cuando se tiene un nivel de censura de hasta el 50% y es más eficiente para niveles de censura mayores, al menos para tamaños de muestras iguales a 3 y 5. Adicionalmente, estudiaron el efecto que tiene el aumento de la proporción de censura en la capacidad de detectar cambios en la media del proceso, encontrando que esta capacidad tiende a disminuir gradualmente a medida que la proporción de datos censurados aumenta para los mismos tamaños de muestra, por lo cual se espera que para un tamaño de muestra  $n > 5$  los resultados sean similares.

## 4 Carta $\bar{X}$ con Censura a Intervalo

En este capítulo se procede a desarrollar tres cartas de control para monitorar la media de un proceso bajo normalidad y censura tipo I a intervalo. Luego, es estudiada la potencia de las cartas propuestas y de la carta Shewhart tradicional, como la probabilidad de detectar cambios en la media de la variable medida. Finalmente, se ejemplifican y comparan las cartas de control propuestas para datos simulados donde a partir de una muestra dada se ha introducido corrimientos en la media.

Como se mencionó previamente, Steiner y Mackay (2000) plantean una situación de control con datos censurados en un proceso productivo, donde se están produciendo tomacorrientes e interesa monitorear el diámetro de sus orificios. Para ello se producen dos galgas con diámetros iguales al límite superior e inferior de especificación y se procede a comparar. Si la galga con mayor diámetro entra en el orificio, entonces el diámetro excede el límite de especificación superior. Si la galga con el menor diámetro no entra en el orificio, entonces el diámetro de éste es menor que el de especificación. En este estudio los orificios que fallan son medidos, por tanto las unidades producidas que sean aceptadas por el experimento quedarán censuradas. En este caso, se presenta censura a intervalo, ya que el valor de la variable de interés, diámetro de los orificios de los enchufes, no será conocido exactamente, sólo se sabrá que está entre el límite superior y el inferior de especificación. Cabe notar que en el esquema de monitoreo planteado se presenta censura tipo I, ya que todas las unidades sometidas al estudio son observadas en relación a un valor del diámetro especificado y se espera una alta proporción de censura, es decir, de productos “buenos”.

Bajo la anterior situación, una aplicación directa de la carta  $\bar{X}$  para monitorear la media del proceso no parece ser adecuada, ya que se está ignorando la censura presente en la muestra y por tanto se ve reducida la cantidad de información disponible, lo que podría llevar a concluir erróneamente que el proceso está bajo control cuando realmente no lo está. De acuerdo a esto, se propone implementar una carta que tome en cuenta la censura y logre aprovechar eficientemente la información disponible en la muestra.

En este punto, la idea expuesta por diferentes autores para tratar la falta de información resultante de la censura, consiste en reemplazar cada observación censurada por un valor cercano al verdadero valor de la variable de interés. Para ello, existen varios métodos en la literatura, como el Valor Esperado Condicional (CEV) implementado en varias de las referencias citadas en el Capítulo 3 de esta tesis, donde para el caso de censura a intervalo, cada observación censurada será reemplazada por el valor esperado de la variable de interés dado que ésta se encuentra entre el límite inferior y superior del intervalo. Otro método



posible, consiste en reemplazar por el límite superior de censura y/o el mencionado por Hsu et. al (2006) donde se propone imputar las observaciones censuradas con el valor de una uniforme en el intervalo de censura.

A continuación se propone la construcción de tres cartas de control para la media con datos censurados a intervalo, mediante diferentes métodos de imputación como los son: el Valor Esperado Condicional (CEV), el límite superior del intervalo y con el valor aleatorio de una uniforme.

## 4.1. Carta de Control CEV- $\bar{X}$ con censura a intervalo

Siguiendo la denominación usada por Steiner y Mackay (2000), la finalidad de la carta de control CEV- $\bar{X}$  es monitorear y detectar cambios en la media del proceso. En este caso se permitirá identificar cambios en el diámetro promedio de los orificios en los tomacorrientes, el cual se desea que no aumente ni disminuya significativamente. Por esta razón, la carta de control CEV- $\bar{X}$  tendrá tanto límite inferior como límite superior de control.

Considerando el procedimiento desarrollado por Steiner y Mackay (2000) para construir la carta CEV  $\bar{X}$  con datos censurados a derecha, se procede a implementar una carta similar para el caso de censura a intervalo. Sea  $X$  la variable que representa el diámetro de los orificios en los enchufes y suponiendo que el proceso está bajo control, dicha variable tiene una distribución normal con media  $\mu$  y desviación estándar  $\sigma$ . La idea es reemplazar cada valor censurado en la muestra por el CEV, para ello se define un intervalo de censura  $[a, b]$ , tal que  $a \leq X \leq b$ , donde  $a$  es el límite inferior y  $b$  el límite superior de control de censura. Bajo estas condiciones la probabilidad de censuramiento está dada por:

$$P_{a,b} = F(b) - F(a) = \phi\left(\frac{a - \mu}{\sigma}\right) - \phi\left(\frac{b - \mu}{\sigma}\right), \quad (4-1)$$

donde  $\phi(\cdot)$  es la función de densidad acumulada normal estándar. De acuerdo a esto, para  $X$  se tiene que:

$$F^*(x) = P(X \leq x | a \leq X \leq b) = \frac{P[(X \leq x) \cap (a \leq X \leq b)]}{P(a \leq X \leq b)}. \quad (4-2)$$

Hay tres casos en el numerador:

$$\text{Si } x < a, \text{ entonces } P[(X \leq x) \cap (a \leq X \leq b)] = 0,$$

$$\text{Si } a \leq x \leq b, \text{ entonces } P[(X \leq x) \cap (a \leq X \leq b)] = P(a \leq X \leq x) = F(x) - F(a),$$

$$\text{Si } x > b, \text{ entonces } P[(X \leq x) \cap (a \leq X \leq b)] = P(a \leq X \leq b) = F(b) - F(a).$$

Luego,

$$F^*(x) = \begin{cases} 0 & \text{Si } x < a, \\ \frac{F(x) - F(a)}{F(b) - F(a)} & \text{Si } a \leq x \leq b, \\ 1 & \text{Si } x > b. \end{cases} \quad (4-3)$$

Derivando la f.d.a respecto a  $x$  se tiene que:

$$f^*(x) = \begin{cases} \frac{f(x)}{F(b) - F(a)} & \text{Si } a \leq x \leq b, \\ 0 & \text{En otro caso.} \end{cases} \quad (4-4)$$

Conociendo  $f^*(x)$  se puede calcular el valor esperado condicional (CEV) como:

$$E[X|a \leq X \leq b] = \int_a^b x f^*(x) dx = \int_a^b x \frac{f(x)}{F(b) - F(a)} dx. \quad (4-5)$$

Como  $X \sim N(\mu, \sigma^2)$ , entonces  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ , así:

$$E[X|a \leq X \leq b] = \frac{1}{F(b) - F(a)} \int_a^b x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx. \quad (4-6)$$

Sustituyendo  $u = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma u + \mu$  y  $du = \frac{dx}{\sigma}$ ,

$$E[X|a \leq X \leq b] = \frac{1}{\Phi(Z_a) - \Phi(Z_b)} \int_{Z_a}^{Z_b} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} (\mu + u\sigma) du$$

$$= \frac{1}{\Phi(Z_a) - \Phi(Z_b)} \left[ \mu \int_{Z_a}^{Z_b} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du + \sigma \int_{Z_a}^{Z_b} u \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du \right],$$

si  $v = \frac{u^2}{2}$  y  $dv = u du$ ,

$$E[X|a \leq X \leq b] = \frac{1}{\Phi(Z_a) - \Phi(Z_b)} \left[ \mu (\Phi(Z_a) - \Phi(Z_b)) + \sigma \int_{\frac{Z_a^2}{2}}^{\frac{Z_b^2}{2}} \frac{1}{\sqrt{2\pi}} \exp\{-v\} dv \right]$$

$$= \mu + \sigma \frac{\frac{1}{\sqrt{2\pi}} \left[ -\exp\left\{-\frac{Z_b^2}{2}\right\} + \exp\left\{-\frac{Z_a^2}{2}\right\} \right]}{\Phi(Z_b) - \Phi(Z_a)}$$

$$\begin{aligned}
&= \mu + \sigma \frac{\left[ -\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{Z_b^2}{2}\right\} + \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{Z_a^2}{2}\right\} \right]}{\Phi(Z_b) - \Phi(Z_a)} \\
&= \mu + \sigma \left[ \frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)} \right] = w_{a,b}.
\end{aligned} \tag{4-7}$$

donde  $Z_a = \frac{a - \mu}{\sigma}$  y  $Z_b = \frac{b - \mu}{\sigma}$ . Finalmente, se define la variable  $W$  para cada observación tal que su valor CEV corresponde a:

$$W = \begin{cases} w_{a,b} & a \leq x \leq b \quad (\text{Censurada}), \\ X & \text{En otro caso} \quad (\text{No Censurada}). \end{cases} \tag{4-8}$$

Los límites superior e inferior de la carta se hallan mediante simulación Monte Carlo y dependen de la proporción de censura  $p$  y del tamaño de subgrupo racional  $n$  usado en el monitoreo del proceso. Para ello se toma en consideración diferentes proporciones de censura, desde  $p = 0.1$  hasta  $p = 0.99$  aumentando en 0.001. Luego, con cada  $p$  se calculan los límites para el intervalo de censura  $[a, b]$  dentro del cual quedan las observaciones censuradas en la muestra. Por último se calcula el valor esperado condicional correspondiente  $w_{a,b}$ , por el cual debe ser reemplazada cada observación que esté dentro del intervalo de censura de acuerdo a la ecuación (4-7).

Se generan entonces,  $m$  subgrupos de tamaño  $n$  de observaciones provenientes de una  $N(0, 1)$ ; luego cada observación cuyo valor esté dentro del intervalo de censura es reemplazada por el correspondiente valor esperado condicional  $w_{a,b}$ ; obteniendo así los  $W_i$  para cada nivel de censura. A continuación se procede a calcular las medias  $\bar{W}_n$  para cada uno de los  $m$  subgrupos de tamaño  $n$  y considerando una probabilidad de falsa alarma de 0.0027, se define el límite inferior de control como el cuantil 0.00135 para la distribución de las medias muestrales ( $\bar{W}_n$ ). Así mismo, el límite superior es hallado como el cuantil 0.99865 para los  $\bar{W}_n$  con un tamaño de subgrupo “muy grande”,  $m \rightarrow \infty$ , que en este caso será de 10000.

$$LCL_1 = q_{0.00135}(\bar{W}_n), \tag{4-9}$$

$$UCL_1 = q_{0.99865}(\bar{W}_n), \tag{4-10}$$

donde  $q_{0.00135}(\bar{W}_n)$  es tal que  $P(\bar{W}_n \leq q_{0.00135}(\bar{W}_n)) = 0.00135$  y  $q_{0.99865}(\bar{W}_n)$  es tal que  $P(\bar{W}_n \leq q_{0.99865}(\bar{W}_n)) = 0.99865$ . Los cuales son aproximados bajo la ley de los grandes números:

$$\lim_{m \rightarrow \infty} \hat{q}_{0.00135}(\bar{W}_{n,1}, \bar{W}_{n,2}, \dots, \bar{W}_{n,m}) = q_{0.00135}(\bar{W}_n), \quad (4-11)$$

$$\lim_{m \rightarrow \infty} \hat{q}_{0.99865}(\bar{W}_{n,1}, \bar{W}_{n,2}, \dots, \bar{W}_{n,m}) = q_{0.99865}(\bar{W}_n), \quad (4-12)$$

donde los  $\{\bar{W}_{n,j}\}_{j=1,\dots,m}$  es una sucesión de medias de muestras de tamaño  $n$  generadas de manera independiente de la distribución de  $W$ . Se aproxima este resultado usando simulación Monte Carlo con  $m = 10000$  subgrupos de tamaño  $n$ , así, el cuantil muestral puede ser definido como el promedio ponderado de estadísticos de orden consecutivos, tal que:

$$Q_{[i](p)} = (1 - \gamma)X_{[j]} + \gamma X_{[j+1]} \quad (4-13)$$

Esta definición de cuantil es descrita por Hyndman y Fan (1996) y está implementada en la función R “quantile”, donde  $1 \leq i \leq 9$ , según el argumento *type* seleccionado para la estimación, que por defecto está fijo en  $type = 7$  y corresponde a (4-13), donde  $(j - m)/n \leq p \leq (j - m + 1)/n$ ,  $X_{[j]}$  es el  $j$ -ésimo estadístico de orden y  $n$  es el tamaño de la muestra. El valor de  $\gamma$  es función de  $j = \lfloor np + m \rfloor$ ,  $g = np + m - j$ , y  $m = 1 - p$ .

Adicionalmente, Gumbel (1939) (citado en Hyndman y Fan (1996)) expone que el cuantil muestral también puede ser obtenido, de manera equivalente, como una interpolación lineal entre los puntos  $(p_{[k]}, X_{[k]})$ , donde  $X_{[k]}$  es el  $k$ -ésimo estadístico de orden y  $p_{[k]} = (k - 1)/(n - 1)$ , en este caso  $p = \text{moda} [F(x_{[k]})]$ .

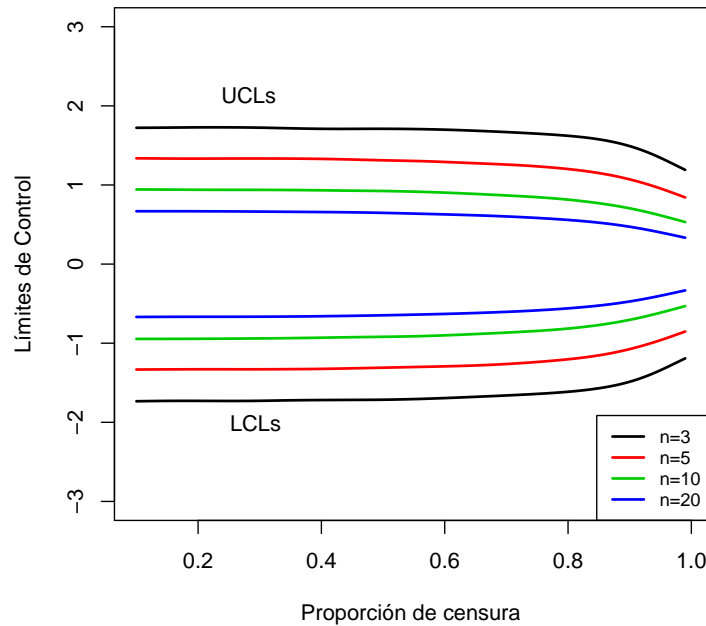
El anterior procedimiento es realizado para los valores de  $n = 3, 5, 10$  y  $20$  y  $m = 10000$  subgrupos, considerando diferentes proporciones de censura, de acuerdo a las cuales se define el intervalo de censura  $[a, b]$ , obteniéndose como resultado la figura 4-1, en la cual se pueden observar los límites simulados para diferentes proporciones de censura  $p$ . Dicha figura muestra los límites calculados luego de aplicar un suavizamiento spline con parámetro  $\lambda = 0.9$ . En la figura 4-1 se logra ver que a medida que aumenta el nivel de censura los límites se estrechan; algo similar sucede a medida que se incrementa el tamaño de subgrupo  $n$ , ya que a medida que  $n$  aumenta, los límites son más estrechos que para tamaños más pequeños.

Los límites de control mostrados en la figura 4-1 son estándar, debido a que fueron calculados asumiendo un proceso bajo control que sigue una distribución normal con media cero y varianza uno. Siendo así, los límites de control requeridos para cualquier proceso con distribución  $N(\mu, \sigma^2)$  se pueden obtener como (Steiner y Mackay, 2000):

$$LCL_{CEV\bar{X}} = LCL_1 * \sigma + \mu, \quad (4-14)$$

$$UCL_{CEV\bar{X}} = UCL_1 * \sigma + \mu, \quad (4-15)$$

donde  $\mu$  y  $\sigma$  son los parámetros del proceso bajo control, con  $LCL_1$  y  $UCL_1$  corresponden a los límites inferior y superior de control estandarizados dados en la figura 4-1.



**Figura 4-1:** UCLs y LCLs Carta CEV- $\bar{X}$ .

Para obtener el valor del límite estandarizado para un tamaño de subgrupo entre los valores calculados se puede hacer una interpolación entre los dos valores de  $n$  conocidos, o realizar todo el proceso de simulación con el  $n$  requerido.

Recapitulando, la construcción de la carta CEV- $\bar{X}$  a intervalo considerando censura Tipo I a intervalo, requiere los siguientes pasos, asumiendo que  $\mu$  y  $\sigma$  son conocidos,

1. Hallar el UCL y LCL mediante la Figura 4-1 dependiendo del tamaño de subgrupo  $n$  y proporción de censura  $p$ .
2. Calcular el valor esperado condicional (CEV) bajo la distribución normal truncada en el intervalo  $[a, b]$  dado en la ecuación (4-7). Luego se reemplaza cada observación censurada por el valor  $w_{a,b}$  hallado.
3. Finalmente, graficar el promedio de cada subgrupo  $\bar{W}_i$  contra los límites calculados. Si hay puntos por fuera de límites se estudian las causas asignables que lo generen, permitiendo así mejorar la calidad del proceso.

## 4.2. Carta de control $\bar{X} - b$ con censura a intervalo.

La carta  $\bar{X} - b$  tiene como objetivo monitorear la media del proceso cuando se presenta censura Tipo I en un intervalo  $[a, b]$  y se imputa el valor de cada observación censurada por el límite superior de dicho intervalo, lo que es muy común en la práctica. La ventaja de esta carta sobre el procedimiento usual o estándar, consiste en que al construirla se tiene en cuenta la censura presente en las observaciones, lo que en la práctica no se considera y usualmente se trabaja como si los valores imputados fuesen los realmente observados.

El procedimiento para construir la carta  $\bar{X} - b$  es similar a la carta CEV  $\bar{X}$ , sólo cambia el método de imputación, donde cada dato censurado será reemplazado por el límite superior de censura. De igual forma, los límites superior e inferior de control de la carta son hallados mediante simulación Monte Carlo, considerando una proporción de censura  $p$  entre 0.1 y 0.99, tamaños de subgrupos  $n = 3, 5, 10$  y 20 y que la variable  $X$  que representa la característica de calidad, tiene una distribución  $N(0, 1)$  para el estudio por simulación.

En este estudio, con el intervalo de censura fijado  $[a, b]$  de acuerdo a la proporción de censura  $p$ , se generan  $m$  subgrupos de tamaño  $n$ , con  $m = 10000$  subgrupos, de una normal estándar; luego se procede a imputar cada observación censurada, es decir dentro de límites, por el valor de  $b$ , generando así la variable  $W$ , y con ella las medias por subgrupo ( $\bar{W}_n$ ), para después, tomando como probabilidad de falsa alarma 0.0027, hallar los límites deseados como los cuantiles 0.00135 y 0.99865 para la distribución de las medias muestrales de la variable  $W$  en cada subgrupo de tamaño  $n$ .

$$W = \begin{cases} b & a \leq X \leq b, \\ X & \text{En otro caso} \end{cases} \quad y \quad X \sim iidN(0, 1). \quad (4-16)$$

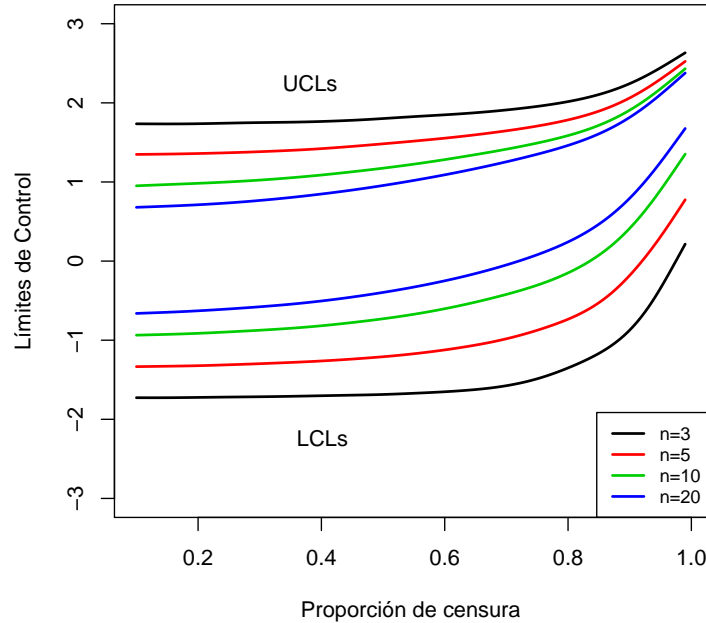
y por tanto los límites bilaterales son calculados como:

$$LCL_2 = q_{0.00135}(\bar{W}_n), \quad (4-17)$$

$$UCL_2 = q_{0.99865}(\bar{W}_n), \quad (4-18)$$

donde  $q_{0.00135}(\bar{W}_n)$  es tal que  $P(\bar{W}_n \leq q_{0.00135}(\bar{W}_n)) = 0,00135$  y  $q_{0.99865}(\bar{W}_n)$  es tal que  $P(\bar{W}_n \leq q_{0.99865}(\bar{W}_n)) = 0.99865$ .

En la Figura 4-2 se tienen los límites de control estándar obtenidos por simulación Monte Carlo para la carta  $\bar{X} - b$ , estos se reportan luego de aplicar un spline con parámetro de suavizamiento  $\lambda = 0.9$ . En dicha figura se observa que los límites dejan de ser simétricos cuando la proporción de censura sobrepasa el 0.2; este comportamiento es esperado debido a que muchas observaciones tomarán el valor del límite superior, por lo cual la distribución



**Figura 4-2:** UCLs y LCLs Carta  $\bar{X} - b$ .

de los  $W$  queda sesgada a derecha. De igual manera que para la carta  $CEV-\bar{X}$ , a medida que el tamaño de subgrupo  $n$  aumenta, los límites se vuelven más estrechos.

En este punto, es claro que los límites obtenidos en la figura 4-2 son estándar, debido al supuesto para la simulación de que  $X \sim N(0, 1)$ . Siendo así, los límites deseados para cualquier proceso  $N(\mu, \sigma^2)$  se pueden obtener como:

$$LCL_{\bar{X}-b} = LCL_2 * \sigma + \mu, \quad (4-19)$$

$$UCL_{\bar{X}-b} = UCL_2 * \sigma + \mu, \quad (4-20)$$

donde  $\mu$  y  $\sigma$  son los parámetros del proceso bajo control, y  $LCL_2$  y  $UCL_2$  son los límites inferior y superior de control estandarizados dados en la figura 4-2.

Para obtener el valor del límite estandarizado para un tamaño de subgrupo entre los valores calculados se puede hacer una interpolación entre los dos valores de  $n$  conocidos, o realizar todo el proceso de simulación con el  $n$  requerido.

Recapitulando, la construcción de la carta  $\bar{X} - b$  a intervalo considerando censura Tipo I a intervalo, implica los siguientes pasos, asumiendo  $\mu$  y  $\sigma$  conocidos,

1. Hallar el UCL y LCL mediante la figura 4-2 dependiendo del tamaño de subgrupo  $n$  y proporción de censura  $p$ .
2. Reemplazar cada observación censurada por el valor del límite superior de censura  $b$  hallado.
3. Finalmente, graficar el promedio de cada subgrupo  $\bar{W}_i$  contra los límites calculados. Si hay puntos por fuera de límites se estudian las causas asignables que lo generen, permitiendo así mejorar la calidad del proceso.

### 4.3. Carta de control $\bar{X} - U$ con censura a intervalo

De manera similar a las cartas propuestas en las secciones anteriores, la carta  $\bar{X} - U$  tiene como finalidad monitorear la media del proceso cuando se presenta censura a intervalo en las observaciones, y su metodología consiste en imputar cada observación censurada por un valor aleatorio de una uniforme dentro del intervalo de censura  $[a, b]$  y calcular los límites como el percentil de las medias de subgrupo luego del proceso de imputación.

Los límites son generados por simulación Monte Carlo, considerando proporciones de censura  $p$  entre 0.1 y 0.99, tamaño de subgrupo  $n = 3, 5, 10, 20$  y que la variable aleatoria  $X \sim N(0, 1)$ . Los tamaños de muestra fueron fijados en los mismos valores usados por Steiner y Mackay (2000) con el fin de comparar los resultados obtenidos por estos autores pero en el caso de censura intervalo planteado en esta tesis, considerando tanto tamaños de muestra pequeños como grandes. Para la simulación, se producen  $m$  subgrupos de tamaño  $n$ , con  $m = 10000$ , luego cada observación dentro de los límites de censura es reemplazada por el valor de una  $U(a, b)$ , produciendo así la variable  $W$  y al calcular las medias de subgrupo se obtienen los  $\bar{W}_n$ , para finalmente definir los límites superior e inferior de control para la carta como cuantiles de la distribución de los  $\bar{W}_n$ , tal que:

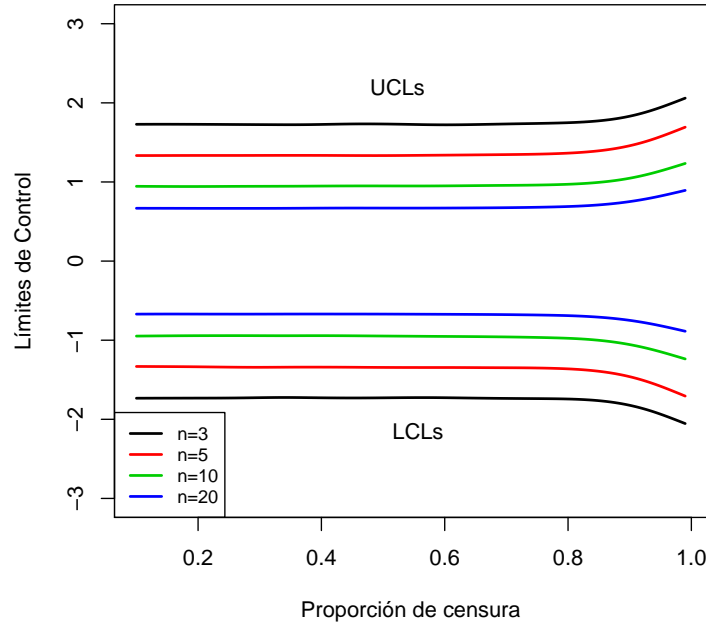
$$LCL_3 = q_{0.00135}(\bar{W}_n), \quad (4-21)$$

$$UCL_3 = q_{0.99865}(\bar{W}_n), \quad (4-22)$$

donde  $q_{0.00135}(\bar{W}_n)$  es tal que  $P(\bar{W}_n \leq q_{0.00135}(\bar{W}_n)) = 0.00135$  y  $q_{0.99865}(\bar{W}_n)$  es tal que  $P(\bar{W}_n \leq q_{0.99865}(\bar{W}_n)) = 0.99865$ .

En la Figura 4-3 se tiene los límites de control estándar obtenidos para la carta  $\bar{X} - U$ , luego de aplicar un spline con parámetro de suavizamiento  $\lambda = 0.9$ , en dicha figura se logra observar que los límites son constantes para cada  $n$  hasta un valor aproximado de  $p = 0.8$ , después del cual se ensanchan para todos los tamaños de muestra; este comportamiento se presenta debido a que a mayor censura existirá mayor variabilidad dentro del intervalo





**Figura 4-3:** UCLs y LCLs Carta  $\bar{X}$  imputando con una  $U(a,b)$ .

de censura, generando límites más amplios para garantizar la proporción de falsa alarma deseada de 0.0027.

Los límites calculados y mostrados en la Figura 4-3 son estándar, debido a que fueron calculados asumiendo un proceso bajo control y sigue una distribución  $N(0, 1)$ . Por tanto, los límites de la carta  $\bar{X} - U$  para cualquier proceso  $N(\mu, \sigma^2)$  se pueden obtener como:

$$LCL_{\bar{X}-U} = LCL_3 * \sigma + \mu, \quad (4-23)$$

$$UCL_{\bar{X}-U} = UCL_3 * \sigma + \mu, \quad (4-24)$$

donde  $\mu$  y  $\sigma$  son los parámetros del proceso bajo control, y  $LCL_3$  y  $UCL_3$  son los límites inferior y superior de control estandarizados dados en la Figura 4-3.

Recapitulando, la construcción de la carta  $\bar{X} - U$  a intervalo considerando censura Tipo I, implica los siguientes pasos, asumiendo  $\mu$  y  $\sigma$  conocidos,

1. Hallar el UCL y LCL mediante la Figura 4-3 dependiendo del tamaño de subgrupo  $n$  y proporción de censura  $p$ .
2. Reemplazar cada observación censurada por el valor aleatorio de una uniforme en el intervalo de censura  $[a, b]$ .

3. Finalmente, graficar el promedio de cada subgrupo  $\bar{W}_i$  contra los límites calculados. Si hay puntos por fuera de límites se estudian las causas asignables que lo generen, permitiendo así mejorar la calidad del proceso.

#### 4.4. Potencia para las cartas $\bar{X}$ propuestas

El desempeño de las cartas de control es medido tradicionalmente en la literatura usando la longitud media de corrida (ARL) como primera medida a considerar (Teoh et. al, 2012), pero la dependencia exclusiva al ARL ha sido criticada por autores como Das (2009). Las críticas se fundamentan en que distribución de la longitud de corrida o RL es muy sesgada a derecha para el proceso bajo control o cuando el cambio en la media del proceso es pequeño; mientras que la desviación estándar de la longitud de corrida, el SDRL, es bastante grande (Das, 2009). Debido a lo cual el ARL no sería la medida adecuada para determinar el rendimiento de una carta, pues pueden obtener conclusiones engañosas basadas en esta medida. Por tanto. En tal caso, Das (2009) recomienda caracterizar la distribución de longitud de corrida por la longitud de ejecución Mediana (LMR) y la desviación estándar de la longitud de ejecución (SDRL).

De acuerdo a lo anterior y tomando en cuenta la propuesta desarrollada por Steiner y Mackay (2000) para medir la potencia de la cartas  $\bar{X}$  mediante la probabilidad de detectar cambios en la media del proceso, se decide usar esta metodología, pues presentó buenos resultados para el caso de censura a derecha en el trabajo realizado por dichos autores.

Por tanto, bajo censura a intervalo en las observaciones y distribución normal, se procede a medir la potencia de las cartas propuestas como la probabilidad de detectar cambios en la media del proceso, tanto crecimientos o decrecimientos, para diferentes proporciones de censura y tamaños de muestra. En este caso se estudiarán proporciones del 50 %, 90 % y 95 % con el fin observar el comportamiento de las cartas cuando la censura es alta, que es lo esperado, según el origen del fenómeno de censura previamente ilustrado. Dichas proporciones de censura son fijadas en las mismas consideradas por Steiner y Mackay (2000) bajo censura a derecha tipo I, con la finalidad de comparar resultados para el caso de censura a intervalo.

Siendo así, considerando el procedimiento desarrollado en Steiner y Mackay (2000), pero para el caso de censura a intervalo, son generados  $m = 10000$  subgrupos de tamaño  $n = 3, 5, 10$  y  $20$  provenientes de una distribución normal estándar con un pequeño cambio en la media medido en unidades de desviación estándar; generando así variaciones en el intervalo  $[-3, 3]$  con cambio de  $0.01$ .

Luego, en la carta CEV- $\bar{X}$  cada observación censurada es reemplazada por el valor esperado condicional, mientras que en la carta  $\bar{X} - b$  es reemplazada por el valor del límite superior de censura  $b$  y por un valor aleatorio de una  $U(a, b)$  en la carta  $\bar{X} - U$ . Obteniéndose así los  $\bar{W}_n$  en cada caso, como el promedio de cada subgrupo de tamaño  $n$ , estos son comparados

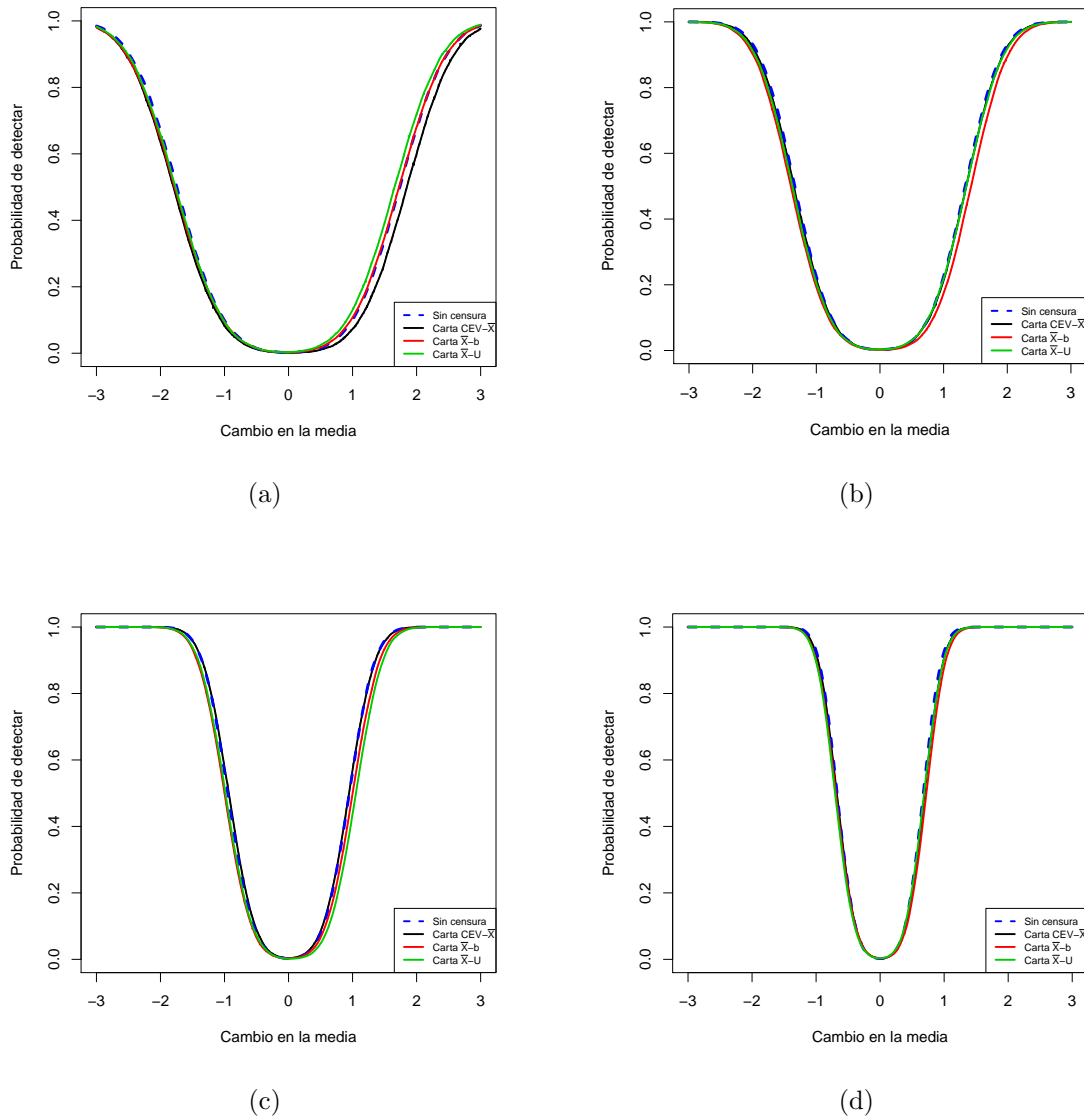
n	p	Carta CEV- $\bar{X}$		Carta $\bar{X}$ -b		Carta $\bar{X}$ -U	
		LCL	UCL	LCL	UCL	LCL	UCL
3	0.5	-1.778	1.828	-1.709	1.769	-1.747	1.637
	0.9	-1.562	1.564	-1.015	2.232	-1.702	1.808
	0.95	-1.564	1.494	-0.748	2.378	-1.961	1.808
5	0.5	-1.311	1.313	-1.228	1.520	-1.338	1.319
	0.9	-1.072	1.122	0.005	2.034	-1.421	1.443
	0.95	-0.996	0.993	0.166	2.264	-1.507	1.568
10	0.5	-0.888	0.897	-0.735	1.177	-0.941	1.007
	0.9	-0.745	0.678	0.451	1.870	-1.062	1.026
	0.95	-0.621	0.643	0.716	2.130	-1.091	1.083
20	0.5	-0.636	0.654	-0.375	0.955	-0.663	0.649
	0.9	-0.522	0.476	0.765	1.786	-0.710	0.753
	0.95	-0.385	0.388	1.177	2.068	-0.815	0.845

**Tabla 4-1:** Límites de control, sin aplicar suavizamiento spline, por carta para proporciones de censura  $p = 0.5, 0.9$  y  $0.95$  y tamaños de muestra  $n = 3, 10$  y  $20$ .

contra los límites inferior y superior de control obtenidos en las Figuras 4-1, 4-2 y 4-3, de acuerdo la carta usada, tamaño de subgrupo y proporción de censura considerada; dichos límites pueden ser observados en la tabla 4-1, los cuales fueron tomados sin spline, pues se desea trabajar con los valores reales obtenidos en la simulación para estudiar su potencia. Finalmente, se cuentan los puntos por fuera de límite y es hallada la probabilidad de detectar cambios en la media del proceso dado que efectivamente se presentó una variación, es decir, se calcula  $P(LCL \leq \bar{W}_n \leq UCL)$ . Adicionalmente, se calcula la potencia de la carta  $\bar{X}$  Shewhart tradicional cuando no hay censura con una probabilidad de falsa alarma de 0.0027, esto con el fin de comparar con los casos con censura.

Las Figuras 4-4, 4-5 y 4-6 muestran las curvas de potencia obtenidas al aplicar el procedimiento de simulación repetido iterativamente considerando el corrimiento en la media con las tres proporciones de censura  $p = 0.5, 0.9$  y  $0.95$  y los tamaños de subgrupo  $n = 3, 5, 10$  y  $20$ . En éstas se puede observar cómo a medida que aumenta el tamaño de subgrupo  $n$ , aumenta la potencia de las cartas en detectar cambios. También a medida que aumenta la proporción de censura, visto de una figura a otra, la potencia de las cartas disminuye, pues se tendría menos información muestral sobre la media del proceso al aumentar la censura.

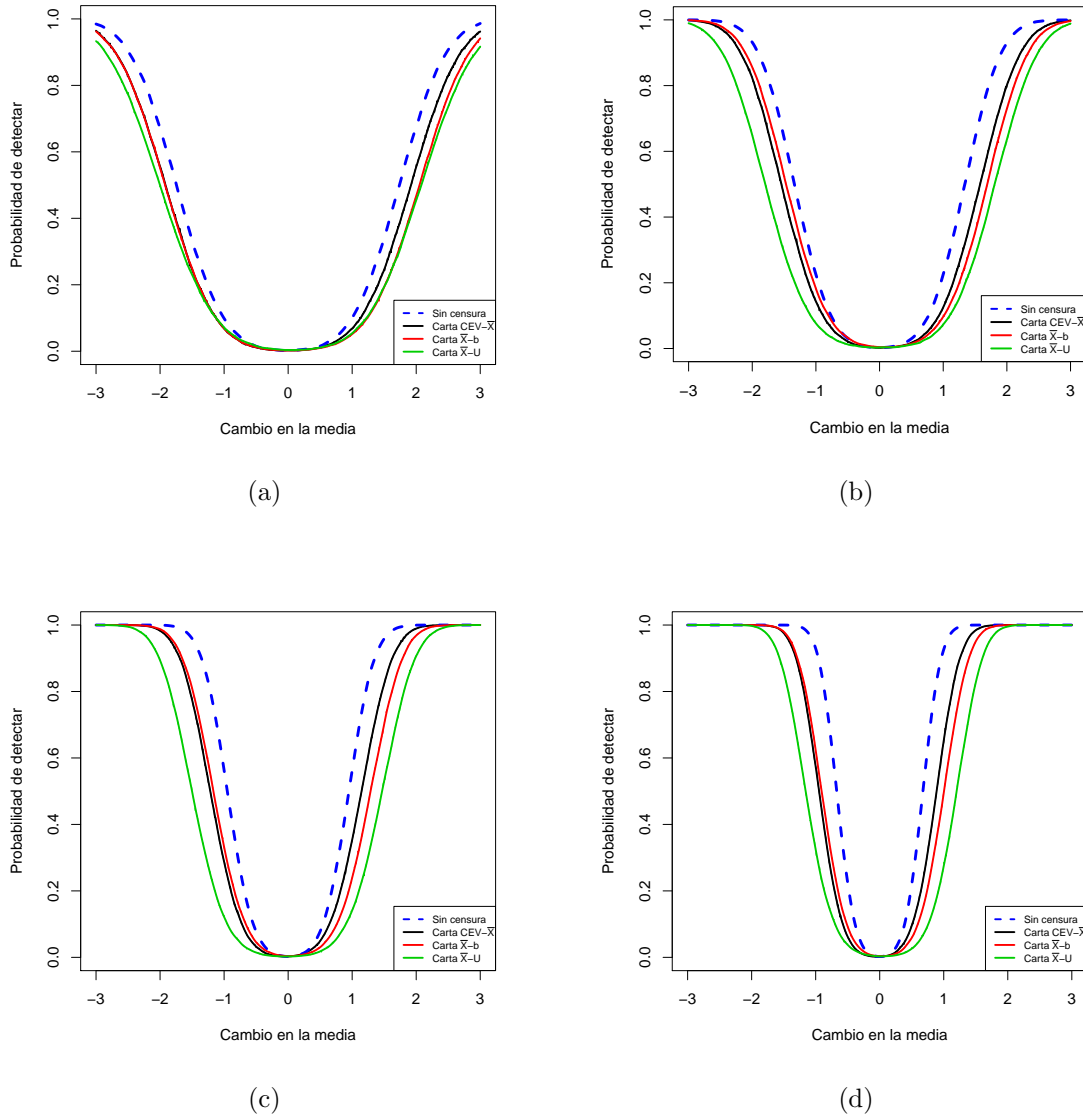
Haciendo un comparativo entre las tres propuestas, se tiene que para un nivel de censura del 50% y una probabilidad de falsa alarma de 0.0027, las tres cartas muestran una potencia similar en la detección de cambios en la media, incluso similar a una carta tipo Shewhart cuando no hay censura en los datos, representada en las Figuras 4-4, 4-5 y 4-6 con la línea



**Figura 4-4:** Potencias comparativas entre cartas propuestas con  $p = 0.5$ . (a)  $n = 3$ ; (b)  $n = 5$ ; (c)  $n = 10$ ; (d)  $n = 20$ .

discontinua. Para una proporción del 90% de censura Figura 4-5 se puede observar que a medida que aumenta  $n$  la potencia de las cartas difiere; para todos los tamaños de muestra la potencia de las cartas  $CEV-\bar{X}$  y  $\bar{X} - b$  es muy similar para decrecimientos en la media, siendo ligeramente mayor la potencia de la segunda, mientras que cuando la media crece la potencia de la carta  $\bar{X} - b$  es claramente menor respecto a la  $CEV-\bar{X}$ . Para dicha proporción de censura del 90%, en todos los casos la potencia de la carta  $\bar{X} - U$  es menor respecto a las otras dos.

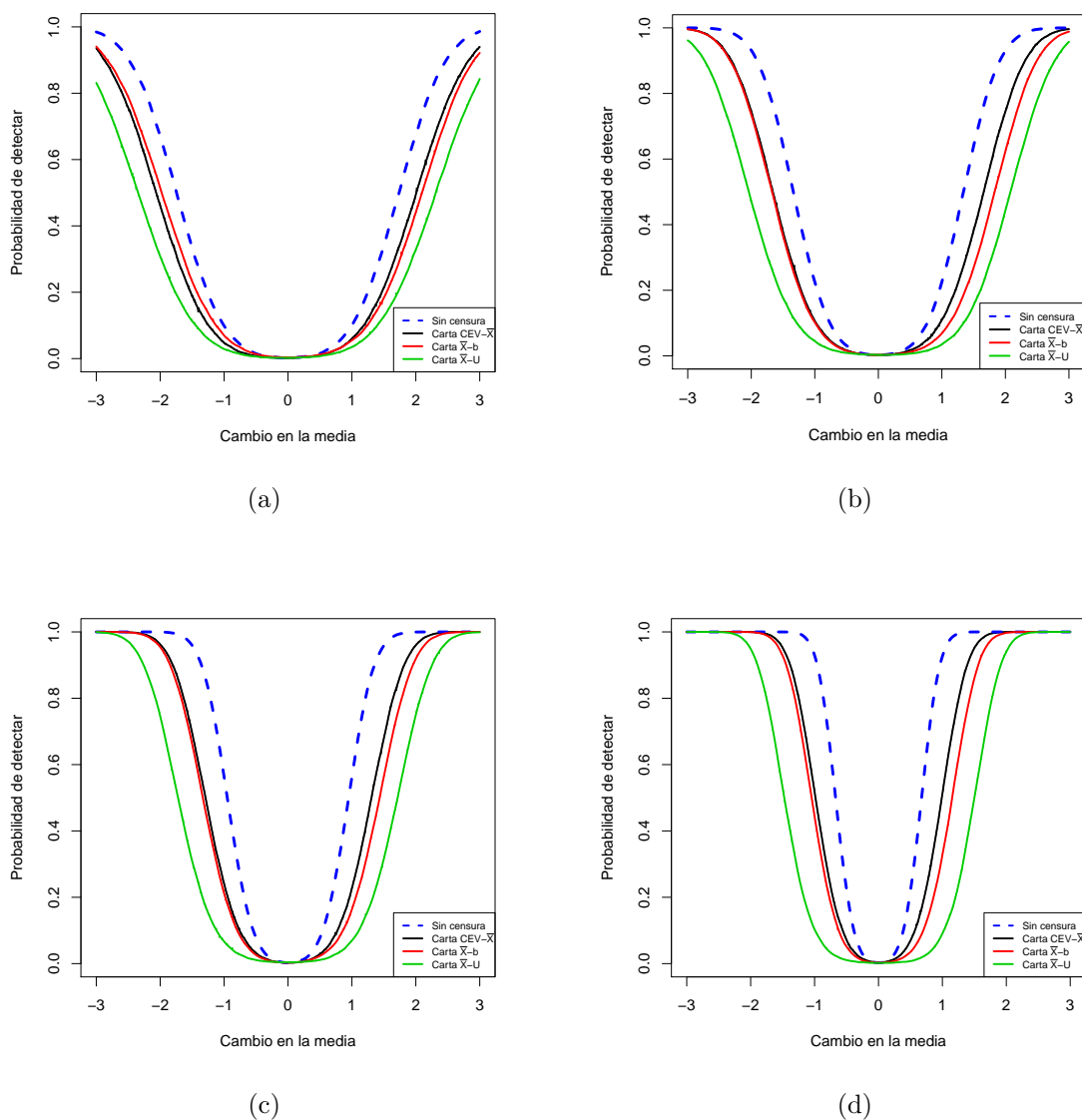
Similarmente, para una proporción del 95% de censura, la Figura 4-6 muestra que la dife-



**Figura 4-5:** Potencias comparativas entre cartas propuestas con  $p = 0.9$ . (a)  $n = 3$ ; (b)  $n = 5$ ; (c)  $n = 10$ ; (d)  $n = 20$ .

rencia en la potencia de cada carta se vuelve más evidente. Para todos los  $n$  considerados la potencia de la carta  $\bar{X} - U$  es menor, convirtiéndose en la menos indicada de las tres cuando se tiene alta censura. En este mismo caso, la potencia de la carta  $CEV - \bar{X}$  es mayor respecto a la carta  $\bar{X} - b$  cuando la media del proceso aumenta, siendo aún mayor esta diferencia a medida que  $n$  aumenta; pero si la media disminuye la detección de cambios es similar entre estas dos cartas, excepto para un tamaño  $n = 20$  donde es notorio un mayor valor en la probabilidad de detección para la carta  $CEV - \bar{X}$ .

En conclusión, de acuerdo a los resultados observados, se prefiere la carta  $CEV - \bar{X}$  para



**Figura 4-6:** Potencias comparativas entre cartas propuestas con  $p = 0.95$ . (a)  $n = 3$ ; (b)  $n = 5$ ; (c)  $n = 10$ ; (d)  $n = 20$ .

monitorear la media del proceso cuando hay censura a intervalo en los datos, pudiéndose hacer excepciones en el caso que se tengan muestras pequeñas, censuras altas y se desea detectar decrecimientos en la media, donde la carta  $\bar{X} - b$  es más potente, pues de acuerdo a las Figuras 4.5(a), 4.6(a), 4.5(b) y 4.6(b) su curva de potencia está ligeramente por encima de la correspondiente a la carta  $CEV - \bar{X}$ , y dicha diferencia puede representar en términos del proceso que se está monitoreando pérdidas económicas, sobre todo en lotes de gran tamaño si no se selecciona la carta más adecuada.

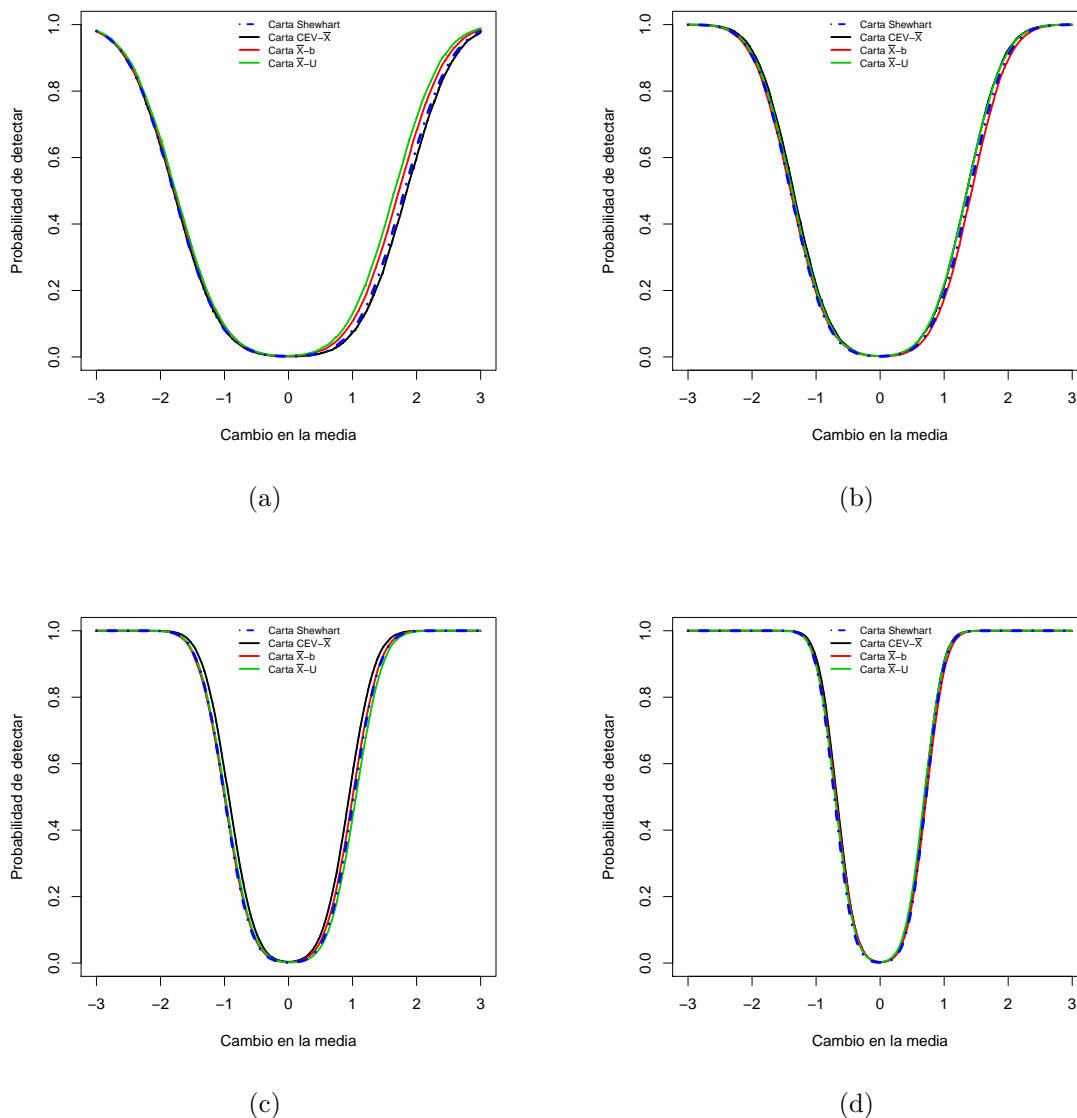
## 4.5. Potencia comparativa de las cartas $\bar{X}$ propuestas Vs. la carta tipo Shewhart tradicional

En la práctica, la carta tipo Shewhart es la más usada para monitorear la media del proceso, por tanto, resulta de interés comparar la potencia de ésta con las tres cartas propuestas cuando hay censura a intervalo en las observaciones. Para poder realizar la comparación todas las cartas deben tener la misma probabilidad de falsa alarma cuando el proceso esté en control, la cual se fija en 0.0027. La carta  $\bar{X}$  tradicional con límites  $\bar{\bar{X}} \pm 3\hat{\sigma}/\sqrt{n}$  ignora la censura y la asimetría presente en las observaciones censuradas. Luego, si se usan los límites  $3\sigma$  la probabilidad de falsa alarma será diferente a la deseada, puesto que el 99,73% de las observaciones no estarán a  $3\sigma$  de la media. Por tanto se debe buscar un valor de  $k$  diferente de 3 que sí satisfaga el nivel de falsa alarma deseado.

n	p	k	LCL	UCL
3	0.50	3.38	-1.78	1.80
	0.90	7.79	-1.73	1.73
	0.95	13.18	-1.73	1.73
5	0.50	3.20	-1.36	1.35
	0.90	6.79	-1.34	1.33
	0.95	11.78	-1.34	1.34
10	0.50	3.14	-0.95	0.95
	0.90	5.62	-0.95	0.96
	0.95	9.09	-0.95	0.95
20	0.50	3.11	-0.67	0.67
	0.90	4.93	-0.67	0.67
	0.95	7.14	-0.67	0.67

**Tabla 4-2:** Límites de control simulados para la carta Shewhart ignorando censura.

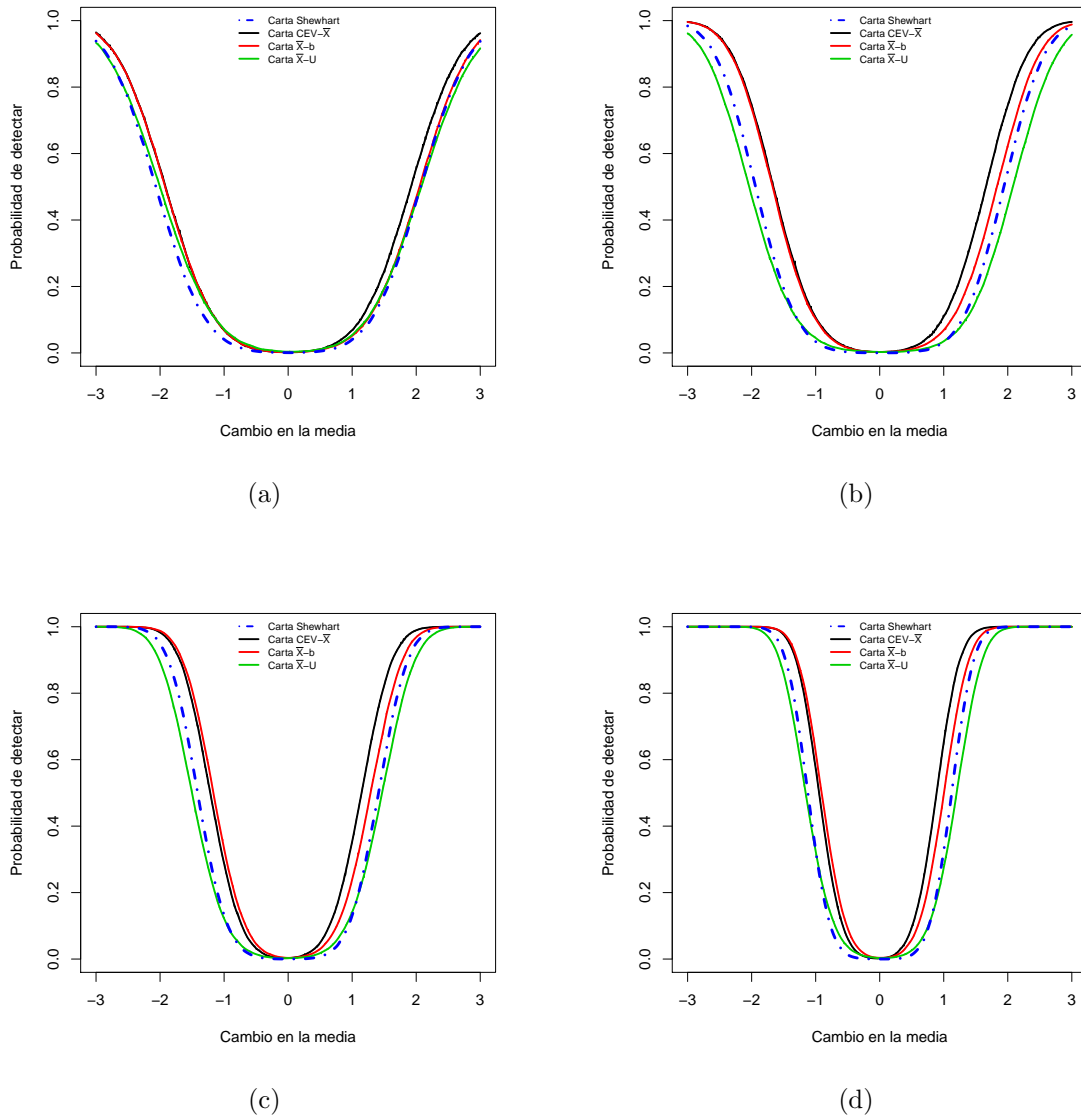
Para construir la carta tipo Shewhart tradicional se procede a generar  $m = 10000$  subgrupos de tamaño  $n = 3, 5, 10$  y  $20$  de una normal estándar, se define el intervalo de censura  $[a, b]$  de acuerdo a las proporciones de censura del 50, 90 y 95 %, luego cada observación dentro del intervalo es tratada como no censurada y toma el valor del punto medio del intervalo, lo que es una práctica común en procesos reales con este tipo de censura y este coincide con el valor esperado condicional. Los límites son calculados como  $\mu \pm k\sigma/\sqrt{n}$ , donde el valor de  $k$  se busca tal que garantice la probabilidad de falsa alarma deseada, es decir, que  $P(|\bar{X} - \mu| > k\sigma/\sqrt{n}) = 0.0027$ , con tamaño de subgrupo  $m \rightarrow \infty$ . De acuerdo a lo anterior, se obtuvieron los valores de  $k$  y límites superior e inferior para la carta tipo Shewhart con los tamaños de subgrupo y proporciones de censura consideradas, los cuales se pueden ver en el Tabla 4-2.



**Figura 4-7:** Potencia comparativa entre la carta Shewhart y las propuestas con  $p = 0.5$ . (a)  $n = 3$ ; (b)  $n = 5$ ; (c)  $n = 10$ ; (d)  $n = 20$ .

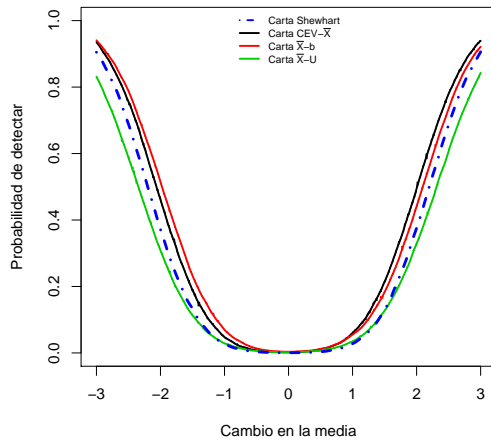
En la figura 4-7 se observa que para proporciones de censura del 50 % la carta tipo Shewhart posee un potencia casi igual a las cartas propuestas y dicha potencia va siendo más idéntica a medida que aumenta  $n$ , por lo que cualquiera de estas cartas podría ser implementada con dicha proporción de censura. Cuando se tiene alta censura, del 90 y 95 %, las Figuras 4-8 y 4-9 muestran que la potencia en la detección de cambios en la media para la carta tipo Shewhart es menor en todos los casos que para la carta propuesta CEV- $\bar{X}$ ; siendo mayor dicha diferencia para un 95 % de censura, reflejando que a mayor censura menos potente es la carta Shewhart para detectar.



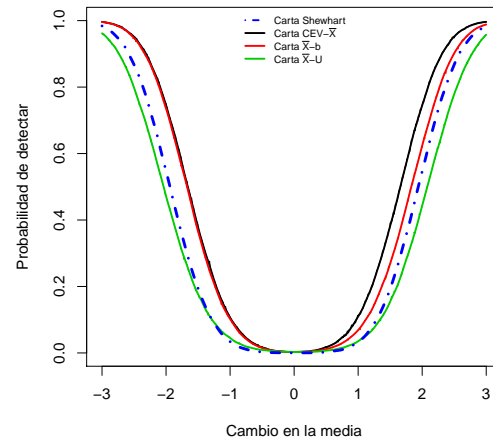


**Figura 4-8:** Potencia comparativa entre la carta Shewhart y las propuestas con  $p = 0.9$ . (a)  $n = 3$ ; (b)  $n = 5$ ; (c)  $n = 10$ ; (d)  $n = 20$ .

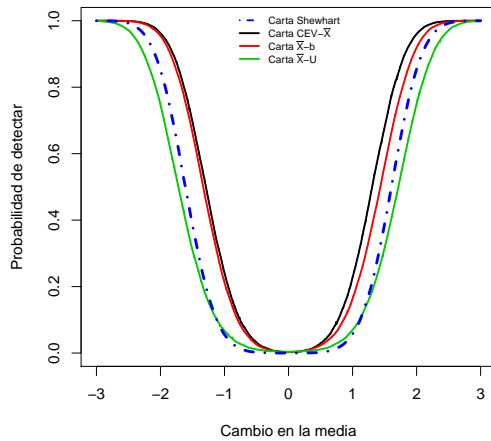
Las figuras 4-7, 4-8 y 4-9 muestran que la potencia de la carta  $\bar{X} - b$  es también mayor que la obtenida para la carta Shewhart tradicional, por lo que su implementación mejoraría en cierta medida la potencia en monitoreo de la media del proceso. Respecto a la carta  $\bar{X} - U$ , la potencia de la carta Shewhart es similar cuando se tiene una muestra pequeña de  $n = 3$  y censura del 95%, pero en todos los demás casos con alta censura gana en potencia la carta tipo Shewhart, reflejando que es preferible monitorear con la media de un proceso con censura a intervalo de la manera tradicional que implementar una carta donde se impute dicha censura con el valor de una uniforme. Este último resultado no resulta extraño, pues



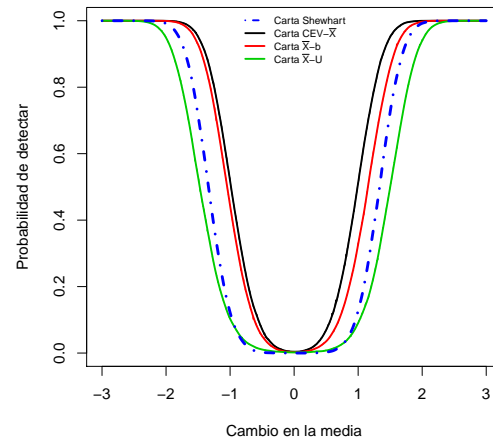
(a)



(b)



(c)



(d)

**Figura 4-9:** Potencia comparativa entre la carta Shewhart y las propuestas con  $p = 0.95$ . (a)  $n = 3$ ; (b)  $n = 5$ ; (c)  $n = 10$ ; (d)  $n = 20$ .

en otras áreas donde se ha imputado la censura con el valor de una uniforme se ha visto su bajo rendimiento respecto a otras metodologías (Hsu, et al., 2006).

En conclusión, las cartas  $CEV-\bar{X}$  y  $\bar{X} - b$  son más potentes que la carta tipo Shewhart tradicional en la detección de corrimientos en la media cuando se tiene alta censura (90 y 95%) para todos los tamaños de muestra; prefiriéndose la carta  $CEV-\bar{X}$  como la que posee mayor potencia en la detección tanto de crecimientos como de decrecimientos en la media del proceso cuando se tiene alta censura y tamaños de muestra moderados y grandes.

## 4.6. Caso simulado

A manera de ejemplo y con el fin de observar el desempeño de las cartas propuestas, se procede a calcular los límites de control para cada una de éstas considerando distribución normal con media  $\mu = 10$  y desviación estándar  $\sigma = 2.5$  y una proporción de censura  $p = 0.95$ . En este caso se considera que los parámetros son conocidos ya que se está en fase II de monitoreo. En el caso de la carta  $CEV-\bar{X}$ , la forma de hallar los límites con parámetros conocidos, está dada en las ecuaciones (4-14) y (4-15), donde  $LCL_1$  y  $UCL_1$  son los límites estándar dados en la Figura 4-1 y en la tabla 4-1 para la proporción de censura  $p = 0.95$  y  $n = 5$  considerados en la simulación, y por tanto,

$$LCL_{CEV-\bar{X}} = -0.9962338 \times 2.5 + 10 = 7.509416 \quad (4-25)$$

$$UCL_{CEV-\bar{X}} = 0.9962338 \times 2.5 + 10 = 12.48267 \quad (4-26)$$

De manera similar, para la carta  $\bar{X} - b$  los límites están dados en las ecuaciones (4-19) y (4-20), donde  $LCL_2$  y  $UCL_2$  son los límites estándar obtenidos de la Figura 4-2 y la tabla 4-1, de lo cual se tiene que,

$$LCL_{\bar{X}-b} = 0.1657177 \times 2.5 + 10 = 10.41429 \quad (4-27)$$

$$UCL_{\bar{X}-b} = 2.264077 \times 2.5 + 10 = 15.66019 \quad (4-28)$$

Para la carta  $\bar{X} - U$  los límites están dados en las ecuaciones (4-23) y (4-24), donde  $LCL_3$  y  $UCL_3$  son los límites estándar tomados de la Figura 4-3 y la tabla 4-1 obteniéndose,

$$LCL_{\bar{X}-U} = -1.507289 \times 2.5 + 10 = 6.231778 \quad (4-29)$$

$$UCL_{\bar{X}-U} = 1.567946 \times 2.5 + 10 = 13.91987 \quad (4-30)$$

Luego, se procede a simular una muestra inicial de  $m_1 = 30$  subgrupos de tamaño  $n = 5$  proveniente de una distribución normal con media  $\mu = 10$ , desviación estándar  $\sigma = 2.5$  y una proporción de censura  $p = 0.95$ , de acuerdo a lo cual se define el intervalo de censura  $[a, b]$ , donde en este caso  $a = 5.1$  y  $b = 14.9$  y cada observación dentro de dicho intervalo es imputada por un valor determinado dependiendo de la carta usada. Para el caso de la carta  $CEV-\bar{X}$ , cada observación censurada es reemplazada por el valor esperado condicional determinado por la ecuación (4-7), el cual para este caso es igual a 10. Así mismo, para la carta  $\bar{X} - b$  la censura es imputada con el valor del límite superior del intervalo  $b = 14.9$  y para la carta  $\bar{X} - U$  se imputa la censura con el valor de una uniforme  $U(a, b)$ . Se decidió construir también la carta tipo Shewhart para la misma muestra simulada donde cada observación dentro del intervalo  $[a, b]$  es reemplazada por el punto medio del intervalo, que en este caso es igual a 10, y sus límites de control son calculados como límites  $3\sigma$ , tal que  $\mu \pm 3\sigma/\sqrt{n}$ , pues con datos completos, bajo normalidad y proceso en control se garantizaría la probabilidad de falsa alarma de 0.0027, lo que en este caso no sucedería ya que hay censura;

de acuerdo a lo anterior se obtiene la Figura 4.10(d), donde se puede notar que la carta no detecta puntos por fuera de límites, dejando percibir su baja potencia cuando se tiene una alta censura, de 95 % en este caso.

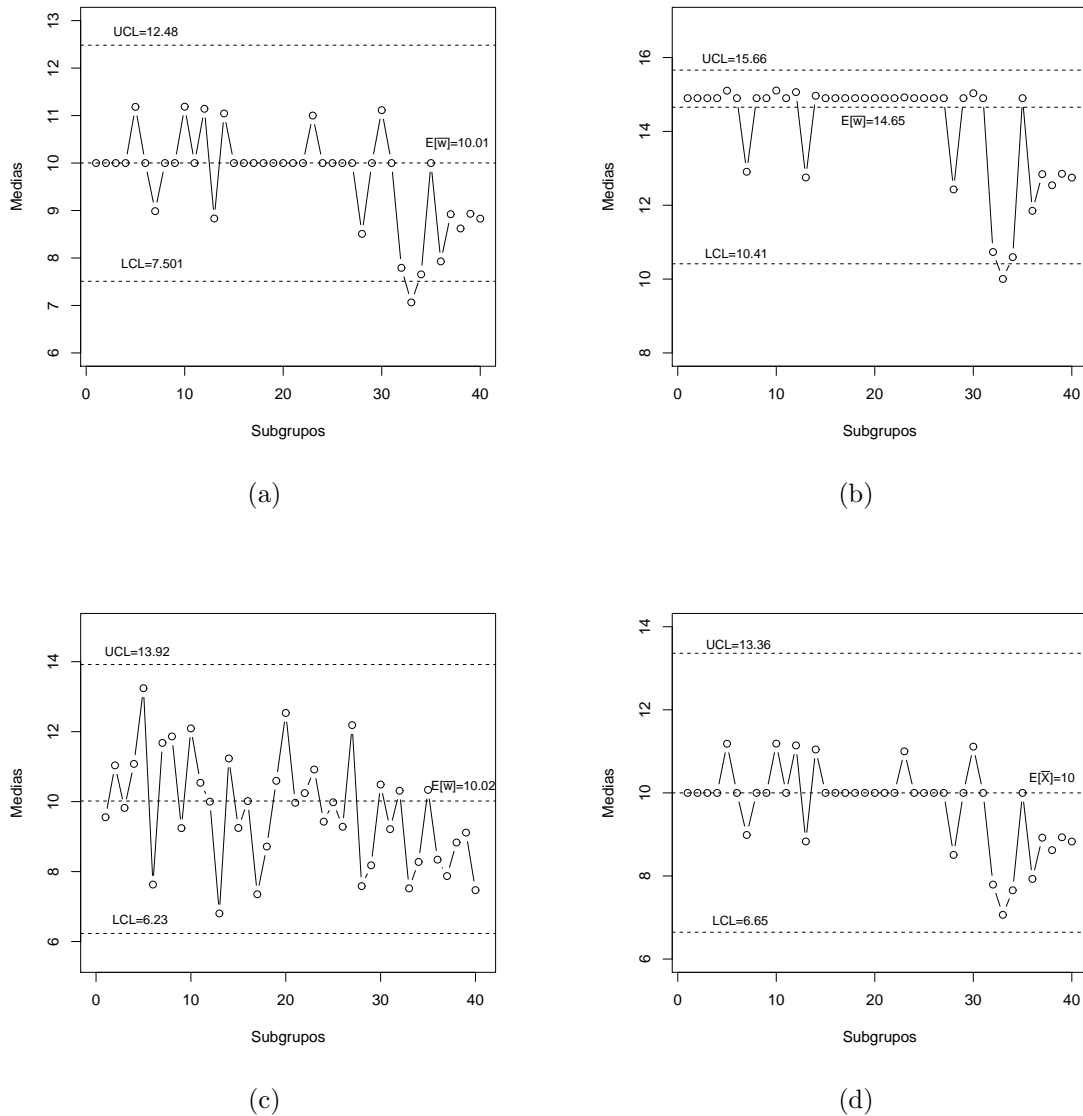
Después, es simulada una muestra adicional de  $m_2 = 10$  subgrupos de tamaño  $n = 5$ , pero esta vez proveniente de una distribución normal con parámetros  $\mu_c = \mu - 1\sigma = 7.5$  y desviación estándar  $\sigma = 2.5$ , es decir con un corrimiento en la media de  $-1\sigma$ , con el fin de estudiar la potencia de las cartas en la detección de este cambio a escalón, donde desde el subgrupo 31 se introduce el cambio en la media. La muestra obtenida fue imputada dentro del intervalo de censura con el valor correspondiente a cada una de las tres cartas y al graficar las 40 medias obtenidas con cada método contra los límites calculados para las cartas propuestas se obtuvo la Figura 4-10, en la cual se grafica el  $E[\bar{W}]$  para cada carta, hallado mediante simulación considerando 10000 subgrupos de tamaño cinco y una censura del 95 %.

La figura 4.10(a) muestra como los límites de la carta CEV- $\bar{X}$  detectan un punto fuera de límites, el cual se presenta en el subgrupo 33 cuya media presenta corrimiento; cabe notar que la carta no muestra señal para las otras medias que provienen de la muestra con variación en la media de  $-1\sigma$  antes mencionada. En la Figura 4.10(b) se puede ver que la carta  $\bar{X} - b$  también detecta señal para la misma media con corrimiento. La carta  $\bar{X} - U$  para la muestra simulada puede ser observada en la Figura 4.10(c), en la cual no se presenta señal para ninguna media de subgrupo, llevando a la conclusión errónea de que el proceso está bajo control cuando realmente no lo está, a menos que se usara reglas de corrida para identificar que ha habido disminución en la media. La carta  $\bar{X} - U$  no fue eficaz para la detección de cambios en la media, reflejando su baja potencia para detectar respecto a las otras dos cartas propuestas para censura a intervalo.

En general, en la Figura 4-10 se observa que los puntos graficados para los 10 últimos subgrupos caen por debajo de la línea media, pero sólo en las cartas CEV- $\bar{X}$  y  $\bar{X} - b$  se detecta señal.

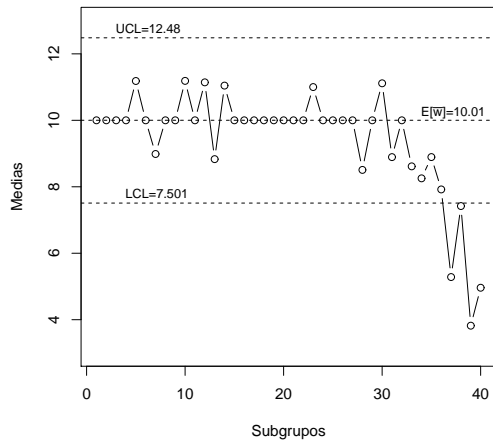
Adicional al caso anterior, donde fueron simulados subgrupos provenientes de una distribución con corrimiento en la media en escalón, es decir que hubo un cambio que se conservó para los siguientes subgrupos, se procede a generar una muestra  $m_3 = 10$  subgrupos, adicional a los 30 en control, del mismo tamaño  $n = 5$  con decrecimientos graduales en la media de  $0.5\sigma$  cada dos subgrupos, generando que los últimos dos presenten un corrimiento de  $2.5\sigma$ , es decir, que a partir de la observación 31 se presenta el cambio; esto con el fin de ejemplificar qué tan rápido es capaz de detectar cada una de las cartas propuestas variaciones constantes en la media. Para la construcción de las cartas se toma en consideración los primeros 30 subgrupos en control y los últimos 10 fuera de control. En todo subgrupo cada observación dentro del intervalo de censura  $[a, b]$  es imputada por el valor correspondiente a cada una de las tres cartas propuestas y los límites de control son los mismos obtenidos para el escenario de simulación anterior, ya que estos son hallados con los subgrupos en control.

En la Figura 4-11 se pueden observar las cartas de control propuestas para los subgrupos

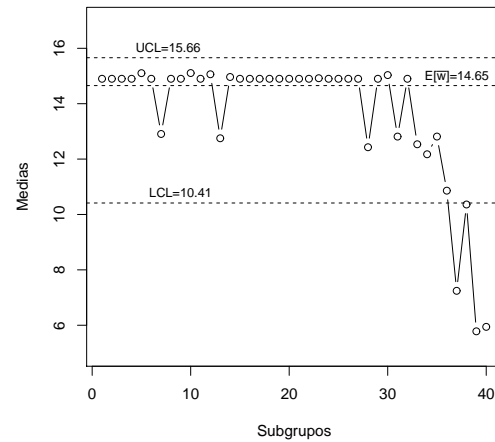


**Figura 4-10:** Cartas propuestas aplicadas a caso simulado con cambio en escalón para la media. (a) Carta  $CEV-\bar{X}$ ; (b) Carta  $\bar{X} - b$ ; (c) Carta  $\bar{X} - U$ ; (d) Carta  $\bar{X}$  Shewhart

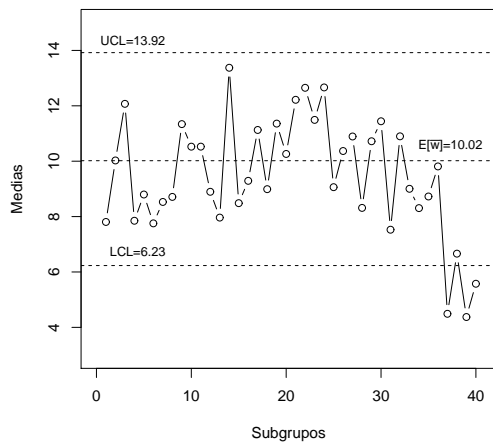
simulados y la carta Shewhart tradicional ignorando censura. En dicha figura se puede ver que para “grandes” corrimientos en la media todas las cartas logran detectar; pero son las cartas  $CEV-\bar{X}$  y  $\bar{X} - b$  que detectan más rápido que el proceso está fuera de control, ya que como se muestra en las Figuras 4.11(a) y 4.11(b), detectan punto por fuera de límites para el subgrupo 37, con un decrecimiento de  $2\sigma$ , e incluso para el subgrupo 36, con un corrimiento de  $1.5\sigma$  ya se está bastante cerca del límite inferior para estas dos cartas, lo cual no se observó en las cartas  $\bar{X} - U$  y  $\bar{X}$  Shewhart. Es de resaltar, que luego de presentarse señal, las cartas  $CEV-\bar{X}$  y  $\bar{X} - b$  no vuelven a indicar otro punto dentro de límites, mientras



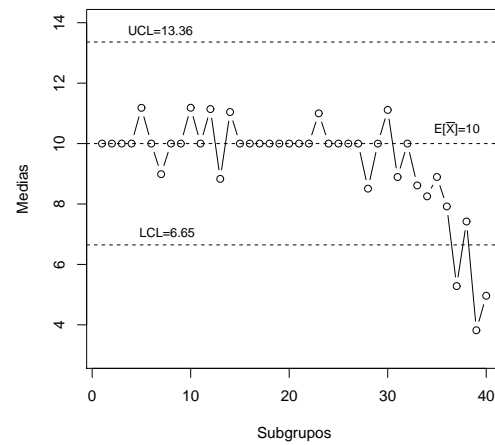
(a)



(b)



(c)



(d)

**Figura 4-11:** Cartas propuestas aplicadas a caso simulado con cambio gradual en la media. (a) Carta CEV- $\bar{X}$ ; (b) Carta  $\bar{X} - b$ ; (c) Carta  $\bar{X} - U$ ; (d) Carta  $\bar{X}$  Shewhart

que en las otras dos cartas,  $\bar{X} - U$  y  $\bar{X}$  Shewhart representadas en las Figuras 4.11(c) y 4.11(d) se tiene como primer punto fuera de límites también el subgrupo 37, pero luego el siguiente lo indican dentro, permitiendo pensar que el proceso sigue en control en el subgrupo 38 y que la señal presente en el anterior pudo ser debida a causas no asignables. Por tanto, estas últimas cartas tardan más en reflejar que el proceso se ha salido de control, ya que el decrecimiento es gradual y no es hasta un decrecimiento en la media de  $2.5\sigma$  que permiten concluirlo.

# 5 Trabajo futuro, conclusiones y recomendaciones

En este capítulo se exponen inicialmente dos alternativas para la estimación de parámetros cuando estos no son conocidos, es decir, se está en la fase I de control y se tiene censura a intervalo tipo I. Dichos métodos son: estimación por máxima verosimilitud y mediante métodos de optimización. Esta propuesta se ubica dentro del trabajo futuro, pues no hace parte de los objetivos de esta tesis y queda pendiente su implementación. Adicionalmente, en este capítulo se enuncian las conclusiones obtenidas con el desarrollo de la tesis, así mismo se proponen recomendaciones y trabajo futuro adicional relacionado con el tema de la censura a intervalo tipo I durante el monitoreo de procesos.

## 5.1. Estimación de parámetros

Las cartas de control propuestas en el Capítulo 4 son para fase II de control, considerando parámetros conocidos, por tanto, en esta sección se describe cómo estimar los parámetros  $\mu$  y  $\sigma$  para el caso de no conocerlos. Aunque el problema de estimación no está dentro del objetivo de la tesis, se quiere aproximar a este para que luego, en un estudio posterior, se pueda determinar cómo afecta tener parámetros estimados. Los métodos descritos en esta parte son: máxima verosimilitud y optimización mediante algoritmos evolutivos.

### 5.1.1. Estimación por máxima verosimilitud

Para las cartas propuestas se considera su aplicación en fase II y por tanto los parámetros de la distribución de la variable de interés son conocidos. Sin embargo, si no fuesen conocidos, y considerando  $m$  muestras históricas del proceso en control, usando máxima verosimilitud pueden ser estimadas la media y la varianza para datos normales censurados a intervalo. Cabe aclarar que sólo será posible estimar por máxima verosimilitud los parámetros del proceso cuando se imputa la censura con el valor esperado condicional, pues en el caso de las otras cartas, el valor de imputación no depende de la distribución de la variable estudiada. Considerando  $r$  como el número de observaciones exactas (no censuradas),  $(n - r)$  el número de observaciones censuradas,  $D$  representa el conjunto de las  $r$  no censuradas y  $X \sim N(\mu, \sigma^2)$  se tiene que:

$$L(\mu, \sigma) = \left\{ \prod_{i \in D} f(x_i) \right\} * \left\{ \prod_{i \notin D} S(x_i) \right\} \quad (5-1)$$

$$L(\mu, \sigma) = \left\{ \prod_{i \in D} \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right\} \right\} * \left\{ \prod_{i \notin D} \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (x_i - \mu)^2 dx \right\} \right\}$$

$$L(\mu, \sigma) = \left\{ \prod_{i \in D} \frac{1}{\sigma} \phi \left( \frac{x_i - \mu}{\sigma} \right) \right\} * \left\{ \prod_{i \notin D} \left[ \Phi \left( \frac{b - \mu}{\sigma} \right) - \Phi \left( \frac{a - \mu}{\sigma} \right) \right] \right\}$$

$$L(\mu, \sigma) = \left\{ \sigma^{-1} \prod_{i \in D} \phi \left( \frac{x_i - \mu}{\sigma} \right) \right\} * \left\{ \Phi \left( \frac{b - \mu}{\sigma} \right) - \Phi \left( \frac{a - \mu}{\sigma} \right) \right\}^{n-r}$$

$$\log L(\mu, \sigma) = -r \log \sigma + \sum_{i \in D} \log \phi \left( \frac{x_i - \mu}{\sigma} \right) + (n-r) \log \left\{ \Phi \left( \frac{b - \mu}{\sigma} \right) - \Phi \left( \frac{a - \mu}{\sigma} \right) \right\} \quad (5-2)$$

La ecuación (5-2) muestra el logaritmo de la verosimilitud bajo normalidad con censura tipo I a intervalo, y a partir de esta se pueden obtener las expresiones para estimar a  $\mu$  y  $\sigma^2$  derivando parcialmente respecto a cada parámetro.

$$m_1 = \frac{\partial L(\theta)}{\partial \mu} \quad m_2 = \frac{\partial L(\theta)}{\partial \sigma^2} \quad (5-3)$$

$$m_1 = \sum_{i \in D} \frac{1}{\phi \left( \frac{x_i - \mu}{\sigma} \right)} \frac{\partial \phi \left( \frac{x_i - \mu}{\sigma} \right)}{\partial \mu} + \frac{n-r}{\left\{ \Phi \left( \frac{b - \mu}{\sigma} \right) - \Phi \left( \frac{a - \mu}{\sigma} \right) \right\}} * \frac{\partial \{ \Phi(Z_b) - \Phi(Z_a) \}}{\partial \mu}, \quad (5-4)$$

donde,

$$\frac{\partial \phi \left( \frac{x_i - \mu}{\sigma} \right)}{\partial \mu} = \left( \frac{x_i - \mu}{\sigma^2} \right) \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \quad (5-5)$$

$$\begin{aligned} \frac{\partial \{ \Phi(Z_b) - \Phi(Z_a) \}}{\partial \mu} &= \frac{\partial}{\partial \mu} \left[ \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} dx \right] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \frac{\partial}{\partial \mu} \left[ \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \right] dx \\ &= \frac{1}{\sqrt{2\pi}\sigma^3} \int_a^b (x - \mu) \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} dx \end{aligned}$$



$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma^3} \left[ \int_a^b x \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\} dx - \mu \int_a^b \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\} dx \right] \\
&= \frac{1}{\sigma^2} \int_a^b x \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\} dx - \frac{\mu}{\sigma^2} \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\} dx.
\end{aligned}$$

De acuerdo a la ecuación (4.1) se tiene que:

$$= \frac{1}{\sigma^2} [F(b) - F(a)] E[X|a \leq x \leq b] - \frac{\mu}{\sigma^2} [F(b) - F(a)],$$

y por ecuación (4-7),

$$= \frac{1}{\sigma^2} [F(b) - F(a)] \left\{ \mu + \sigma \left[ \frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)} \right] \right\} - \frac{\mu}{\sigma^2} [F(b) - F(a)],$$

finalmente,

$$\frac{\partial \{\Phi(Z_b) - \Phi(Z_a)\}}{\partial \mu} = \frac{\phi(Z_a) - \phi(Z_b)}{\sigma} \quad (5-6)$$

reemplazando (5-5) y (5-6) en (5-4):

$$m_1 = \sum_{i \in D} \left( \frac{x_i - \mu}{\sigma^2} \right) + \frac{(n - r) [\phi(Z_a) - \phi(Z_b)]}{\sigma [\Phi(Z_b) - \Phi(Z_a)]}, \quad (5-7)$$

por ecuación (4-8) se tiene que para el caso de la carta CEV- $\bar{X}$ ,

$$\frac{w_{a,b} - \mu}{\sigma^2} = \frac{\phi(Z_a) - \phi(Z_b)}{\sigma [\Phi(Z_b) - \Phi(Z_a)]} \quad (5-8)$$

así,

$$\begin{aligned}
m_1 &= \sum_{i \in D} \left( \frac{w_i - \mu}{\sigma^2} \right) + \sum_{i \notin D} \left( \frac{w_{a,b} - \mu}{\sigma^2} \right) = \sum_{i=1}^n \left( \frac{w_i - \mu}{\sigma^2} \right) = 0 \\
&\quad \sum_{i=1}^n w_i - \sum_{i=1}^n \mu = 0 \\
\hat{\mu}_1 &= \frac{\sum_{i=1}^n w_i}{n} \quad (5-9)
\end{aligned}$$

Ahora se procede a hallar  $m_2 = \frac{\partial \log L(\theta)}{\partial \sigma^2}$ , sea  $\theta = \sigma^2$ ,

$$m_2 = \frac{\partial}{\partial \theta} \left\{ -r \log \theta^{1/2} + \sum_{i \in D} \log \phi \left( \frac{-\mu}{\theta^{1/2}} \right) + (n-r) \log \left[ \Phi \left( \frac{b-\mu}{\theta^{1/2}} \right) - \Phi \left( \frac{a-\mu}{\theta^{1/2}} \right) \right] \right\}$$

$$m_2 = -\frac{r}{2\theta} + \sum_{i \in D} \frac{1}{2\theta^2} (x-\mu)^2 + \frac{n-r}{\Phi(Z_b) - \Phi(Z_a)} \frac{\partial}{\partial \theta} [\Phi(Z_b) - \Phi(Z_a)], \quad (5-10)$$

donde,

$$\begin{aligned} \frac{\partial}{\partial \theta} [\Phi(Z_b) - \Phi(Z_a)] &= \\ &= \frac{\partial}{\partial \theta} \int_a^b \frac{1}{\sqrt{2\pi\theta^{1/2}}} \exp \left\{ -\frac{1}{2\theta} (x-\mu)^2 \right\} dx = \int_a^b \frac{\partial}{\partial \theta} \left[ \frac{1}{\sqrt{2\pi\theta^{1/2}}} \exp \left\{ -\frac{1}{2\theta} (x-\mu)^2 \right\} \right] dx \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b \left\{ \frac{\theta^{1/2} \exp \left\{ -\frac{1}{2\theta} (x-\mu)^2 \right\} \left( \frac{1}{2\theta^2} (x-\mu)^2 \right) - \exp \left\{ -\frac{1}{2\theta} (x-\mu)^2 \right\} \left( \frac{1}{2} \theta^{-1/2} \right)}{\theta} \right\} dx \\ &= \frac{1}{2\theta^{3/2}} \int_a^b \frac{(x-\mu)^2}{\theta} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi\theta^{1/2}}} \exp \left\{ -\frac{1}{2\theta} (x-\mu)^2 \right\} dx - \frac{1}{2\theta} \int_a^b \frac{1}{\sqrt{2\pi\theta^{1/2}}} \exp \left\{ -\frac{1}{2\theta} (x-\mu)^2 \right\} dx \\ &= \frac{1}{2\theta^{3/2}} \int_a^b \frac{(x-\mu)^2}{\theta} \frac{1}{\sqrt{2\pi\theta^{1/2}}} \exp \left\{ -\frac{1}{2\theta} (x-\mu)^2 \right\} dx - \frac{1}{2\theta} [\Phi(Z_b) - \Phi(Z_a)], \end{aligned}$$

sea  $w = \frac{x-\mu}{\theta^{1/2}}$  y  $dw = \theta^{1/2} dx$ , entonces:

$$= \frac{1}{2\theta^{3/2}} \int_a^b \frac{(x-\mu)^2}{\theta} \frac{1}{\sqrt{2\pi\theta^{1/2}}} \exp \left\{ -\frac{1}{2\theta} (x-\mu)^2 \right\} dx = \theta^{1/2} \int_{Z_a}^{Z_b} \frac{1}{\sqrt{2\pi}} w^2 \exp \left\{ -\frac{1}{2} w^2 \right\} dw,$$

sea  $u = w$ ,  $du = dw$ ,  $dv = w \exp \left\{ -\frac{1}{2} w^2 \right\}$  y  $v = -\exp \left\{ -\frac{1}{2} w^2 \right\}$ , integrando se obtiene que:

$$\begin{aligned} \theta^{1/2} \int_{Z_a}^{Z_b} \frac{1}{\sqrt{2\pi}} w^2 \exp \left\{ -\frac{1}{2} w^2 \right\} dw &= \\ &= \frac{\theta^{1/2}}{\sqrt{2\pi}} \left[ -Z_b \exp \left\{ -\frac{1}{2} Z_b^2 \right\} + Z_a \exp \left\{ -\frac{1}{2} Z_a^2 \right\} \right] + \theta^{1/2} \int_{Z_a}^{Z_b} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} w^2 \right\} dw \end{aligned}$$

$$= \frac{\theta^{1/2}}{2\theta^{3/2}} [Z_a\Phi(Z_a) - Z_b\phi(Z_b)] + \frac{\theta^{1/2}}{2\theta^{3/2}} [\Phi(Z_b) - \phi(Z_a)] - \frac{1}{2\theta} [\Phi(Z_b) - \phi(Z_a)],$$

finalmente,

$$\frac{\partial}{\partial\theta} [\Phi(Z_b) - \Phi(Z_a)] = \frac{1}{2\theta} [Z_a\Phi(Z_a) - Z_b\phi(Z_b)], \quad (5-11)$$

reemplazando (5-11) en (5-10):

$$m_2 = -\frac{r}{2\theta} + \sum_{i \in D} \frac{1}{2\theta^2} (x_i - \mu)^2 + \frac{n-r}{\Phi(Z_b) - \Phi(Z_a)} \frac{1}{2\theta} [Z_a\Phi(Z_a) - Z_b\phi(Z_b)], \quad (5-12)$$

de acuerdo a la ecuación (4-8), para la carta CEV- $\bar{X}$  se obtiene que:

$$\frac{w_i - \mu}{\sigma} = \begin{cases} \frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)} & a \leq x \leq b, \\ \frac{x_i - \mu}{\sigma} & \text{En otro caso.} \end{cases}$$

Luego,

$$\begin{aligned} \sum_{i=1}^n \frac{(w_i - \mu)^2}{\sigma^2} &= \sum_{i \in D} \frac{(x_i - \mu)^2}{\sigma^2} + \sum_{i \notin D} \left[ \frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_a) - \Phi(Z_b)} \right]^2 \\ \sum_{i \in D} \frac{(x_i - \mu)^2}{\sigma^2} &= \sum_{i=1}^n \frac{(w_i - \mu)^2}{\sigma^2} - \sum_{i \notin D} \left[ \frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_a) - \Phi(Z_b)} \right]^2 \end{aligned} \quad (5-13)$$

reemplazando (5-13) en (5-12):

$$\begin{aligned} m_2 &= -r + \frac{1}{\sigma^2} \sum_{i=1}^n (w_i - \mu)^2 - \sum_{i \notin D} \left[ \frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_a) - \Phi(Z_b)} \right]^2 + \frac{n-r}{\Phi(Z_b) - \Phi(Z_a)} [Z_a\Phi(Z_a) - Z_b\phi(Z_b)] = 0 \\ \frac{1}{\sigma^2} \sum_{i=1}^n (w_i - \mu)^2 &= r + (n-r) \left\{ \left[ \frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_a) - \Phi(Z_b)} \right]^2 - \frac{Z_a\Phi(Z_a) - Z_b\phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)} \right\} \end{aligned}$$

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^n (w_i - \hat{\mu}_0)^2}{r + (n-r) \left\{ \left[ \frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_a) - \Phi(Z_b)} \right]^2 - \frac{Z_a\phi(Z_a) - Z_b\phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)} \right\}}. \quad (5-14)$$

Basados en las expresiones obtenidas en las ecuaciones (5-9) y (5-14) es posible aplicar un método iterativo propuesto por Sampford y Taylor (1959), el cual consiste en una aplicación del algoritmo de Esperanza-Maximización (EM), con el cual luego de reemplazar cada observación censurada con el valor esperado condicional dado en la ecuación (4-7), se reemplaza inicialmente  $\mu$  y  $\sigma$  por la media y desviación estándar muestral ( $\hat{\mu}_0, \hat{\sigma}_0$ ) ignorando censura, para luego aplicar iterativamente las ecuaciones (5-9) y (5-14) hasta que los valores para  $\mu$  y  $\sigma$  no varíen significativamente. Se puede definir que si dicho cambio es de 0.01 o más pequeño esos valores estimados serán los que finalmente se considerarán como los valores bajo control. Es de resaltar, que también existe en la literatura métodos de optimización que permiten optimizar el logaritmo de la verosimilitud con el fin de encontrar los parámetros del proceso en control, logrando de manera alternativa estimarlos. Tanto el método analítico y el uso de algoritmos de optimización son implementados en la práctica, por lo que cualquiera de los dos puede ser usado. Es de interés estudiar cuál de los dos procedimientos es más eficiente, pero esto queda fuera del alcance de este trabajo.

En el caso de tener una censura del 100 %, es decir que todas las observaciones quedan censuradas, la función de verosimilitud se reduciría a:

$$\begin{aligned}
 L(\mu, \sigma) &= \prod_{i=1}^n S(x_i) = \prod_{i=1}^n \left[ \Phi \left( \frac{b - \mu}{\sigma} \right) - \Phi \left( \frac{a - \mu}{\sigma} \right) \right] \\
 L(\mu, \sigma) &= \left\{ \Phi \left( \frac{b - \mu}{\sigma} \right) - \Phi \left( \frac{a - \mu}{\sigma} \right) \right\}^n \\
 \log L(\mu, \sigma) &= n \log \left\{ \Phi \left( \frac{b - \mu}{\sigma} \right) - \Phi \left( \frac{a - \mu}{\sigma} \right) \right\}. \tag{5-15}
 \end{aligned}$$

La ecuación (5-15) muestra que en la función de verosimilitud no relaciona la variable de control cuando se presenta censura del 100 % y por tanto no hay información para poder estimar los parámetros, generando que no existan estimadores máximo verosímiles. En este caso, es necesario realizar mediciones exactas de la característica de calidad que se desea monitorear en fase I, y poder estimar los parámetros. Cuando se está en fase II, dado que se conocen los parámetros, es posible aplicar el mecanismo de monitoreo en el que se compara con un patrón la variable de interés, lo que genera censura a intervalo. Por tanto, en fase I, no debe haber censura para poder estimar los parámetros del proceso en control y si en fase II de monitoreo se presenta censura, se debe usar una carta de control que considere esta situación, como lo es la carta CEV- $\bar{X}$  para censura a intervalo descrita en la sección 4.1.

### 5.1.2. Estimación por métodos de optimización

Asumiendo que no hay censura al 100 %, si se desconocen los parámetros del proceso  $\mu$  y  $\sigma$  bajo control, estos pueden ser estimados mediante la implementación de algoritmos de

optimización. Por ejemplo, usando algoritmos de evolución diferencial para maximizar el logaritmo de la función de verosimilitud para el caso de censura a intervalo dada en la ecuación (5-2).

El algoritmo de Evolución Diferencial (DE) es un algoritmo de optimización estocástico introducido por Storn y Price en 1996. Dicho algoritmo engloba heurísticas basadas en la selección natural usando operadores como la mutación, cruce, y selección en cada generación, con el fin de mover la población hacia el óptimo global (Ardia, et. al., 2013). El algoritmo de evolución diferencial ha sido aplicado con éxito para resolver problemas de optimización, obteniendo notable desempeño con una amplia variedad de funciones, en términos de precisión, velocidad de convergencia y robustez (Lin, 2013). Para la implementación del algoritmo son seleccionados tres individuos como padres, donde uno es el principal y éste es perturbado con el vector diferencia de los otros dos padres. De esta manera se crean nuevas soluciones del problema, donde mediante el uso de una población secundaria se busca retener las mejores soluciones en cada generación del proceso evolutivo.

Sea  $S \in R^d$  el espacio de búsqueda del problema, donde  $d$  denota la dimensión, el DE involucra una población de NP vectores o soluciones candidatas  $X_{i,g} = \{x_{1i,g}, x_{2i,g}, \dots, x_{di,g}\} \in S, i = 1, 2, \dots, NP$ . Donde cada  $x_{ji,g}$  se compara con una variable de decisión en el problema y  $g$  indica la generación a la cual pertenece el vector.

Para crear la primera generación, se supone un valor inicial óptimo para NP miembros del vector de parámetros, mediante el uso de números aleatorios entre un límite superior e inferior o usando valores iniciales, ambos definidos por el usuario. En cada generación se crea nueva población de miembros  $\{x_i | i = 1, \dots, NP\}$ , con  $i$  igual al índice del vector que compone la población. Luego por cada vector  $X_{i,g}$  de la generación  $g$  se genera un vector mutado  $V_{i,g} = \{v_{1i,g}, v_{2i,g}, \dots, v_{di,g}\}$  mediante el uso de alguna estrategia de mutación. El vector  $V_{i,g}$  puede ser determinado por (Lin, 2013):

$$V_{i,g} = x_{i1,g} + F(x_{i2,g} - x_{i3,g}) \quad (5-16)$$

Donde  $F$  es el factor de escala entre  $[0, 2]$ , el cual escala la diferencia de vectores, la mutación. Luego se genera la operación de cruce o recombinación, la cual se aplica a cada componente del vector mutado  $V_{i,g}$  y el correspondiente vector objetivo  $X_{i,g}$ , generando así el vector de prueba  $U_{i,g} = \{u_{1i,g}, u_{2i,g}, \dots, u_{di,g}\}$ . Tal que:

$$u_{ji,g} = \begin{cases} v_{ji,g} & \text{si } (U_j[0, 1] \leq CR) \quad \text{ó} \quad (j = j_{rand}), \\ x_{ji,g} & \text{en otro caso.} \end{cases} \quad (5-17)$$

Con  $j=1, 2, \dots, d$ . CR es la constante que indica la probabilidad de cruce en el rango  $[0,1]$  y  $j_{rand}$  es un entero aleatorio dentro del rango  $[1, NP]$ , el cual permite asegurar que el vector de prueba sea diferente del vector objetivo correspondiente. Finalmente viene la selección, la

cual compara el mejor valor de prueba  $f(U_{i,g})$  con su correspondiente vector objetivo  $f(X_{i,g})$  de la población actual. En este caso, el mejor vector será el que pasará a formar parte de la población para la siguiente generación.

$$X_{i,g+1} = \begin{cases} U_{i,g} & \text{si } f(U_{i,g}) < f(X_{i,g}), \\ X_{i,g} & \text{en otro caso.} \end{cases} \quad (5-18)$$

La mutación, cruce y selección son repetidas de generación en generación hasta que se satisfaga algún criterio de detención del algoritmo, como el número de iteraciones permitidas, tiempo transcurrido, entre otras. Durante la aplicación de evolución diferencial se puede presentar una temprana convergencia a óptimos locales, sobretodo en funciones complejas de alta dimensión; dicha velocidad de convergencia depende del factor de mutación  $F$  y de acuerdo al valor que este tome. En general, para realizar búsqueda del óptimo global se debe trabajar con un valor de  $F$  grande (Zhang et. al., 2009).

El algoritmo de evolución diferencial (DE) se encuentra implementado en R en la librería `DEoptim` cuya descripción se puede ver en el apéndice A. Similarmente es posible estimar la media y la varianza del proceso bajo control usando la función `optim` en R, descrita en el apéndice B. Como lo plantea Lin (2013) el algoritmo de evolución diferencial tiene un gran desempeño a la hora de optimizar una función objetivo y ha venido siendo usado recientemente por diversos autores en distintas áreas, por lo que resulta un método recomendable de aplicar en el problema de estimación presente en el estudio como trabajo futuro.

p	n	$\hat{\mu}$	$\hat{\mu} - \mu$	$\hat{\sigma}$	$\hat{\sigma} - \sigma$
0.50	3	9.644	-0.356	2.595	0.095
	5	9.7	-0.3	2.584	0.084
	10	9.73	-0.27	2.647	0.147
	20	9.909	-0.091	2.605	0.105
0.90	3	9.693	-0.307	2.492	-0.008
	5	9.526	-0.474	2.433	-0.067
	10	9.843	-0.157	2.611	0.111
	20	10.055	0.055	2.614	0.114
0.95	3	11.384	1.384	1.961	-0.539
	5	9.704	-0.296	2.535	0.035
	10	9.633	-0.367	2.535	0.035
	20	10.121	0.121	2.573	0.073

**Tabla 5-1:** Parámetros estimados mediante algoritmos evolutivos y sus sesgos, con censura  $p = 0.5, 0.9$  y  $0.95$  y tamaños de muestra  $n = 3, 5, 10$  y  $20$ .

Con el fin de ejemplificar los resultados de la estimación mediante el uso del algoritmo de evolución diferencial se recurre a simulación Monte Carlo, y son generados  $m = 30$  subgrupos de tamaño  $n = 3, 5, 10, 20$  con proporciones de censura  $p = 0.5, 0.9, 0.95$ , de una normal con

media  $\mu = 10$  y desviación estándar  $\sigma = 2.5$ , luego cada observación dentro del intervalo de censura  $[a, b]$  es reemplazada por el valor esperado condicional  $w_{a,b}$  y aplicando el función `DEoptim` descrita en el apéndice A sobre la verosimilitud dada en la ecuación (5-2) son estimados los parámetros  $\mu$  y  $\sigma$  del proceso en control. Dichas estimaciones se pueden ver en la tabla 5-1, la cual refleja que a medida que aumenta el tamaño de muestra  $n$  la precisión de las estimaciones de la media es mayor, pues la diferencia entre el valor estimado y la media real de los datos simulados disminuye en valor absoluto. Respecto a la estimación de la desviación estándar, las desviaciones respecto al verdadero no presentan grandes variaciones al aumentar  $n$ . Cuando se tiene alta censura, del 95%, y un tamaño de muestra pequeño,  $n = 3$ , las estimaciones son las menos precisas, pues en este caso se tiene la menor cantidad de información sobre la variabilidad del proceso de todos los considerados. En general, las estimaciones mediante el uso de este algoritmo son cercanas al valor real y mejoran a medida que aumenta  $n$ .

## 5.2. Conclusiones

- Con alta censura en la muestra, más del 90%, la potencia de la carta  $CEV-\bar{X}$  es la mayor de todas y dicha diferencia es más notoria a medida que aumenta el tamaño de subgrupo racional  $n$ . Con alta censura, la carta  $\bar{X} - b$  presenta un potencia similar a la  $CEV-\bar{X}$  a la hora detectar decrecimientos en la media, pero se queda rezagada en la detección de crecimientos en el valor medio de la característica de calidad que se está monitoreando. La carta  $\bar{X} - U$  es la menos recomendable, pues su potencia es significativamente menor respecto a las otras cartas en el caso de alta censura.
- La potencia de las cartas  $CEV-\bar{X}$  y  $\bar{X} - b$  es mayor respecto a la carta Shewhart cuando se tiene alta censura y por tanto implementar dichas cartas mejora la eficiencia del monitoreo.
- En el esquema planteado en este trabajo se prefiere una alta censura, pues las observaciones estarían dentro de las medidas deseadas, por lo que se hace necesario determinar tamaños de muestra más grandes que en el caso de censura baja, pues así es posible obtener mayor información sobre la variabilidad del proceso.
- Las cartas de control propuestas no sólo permiten ver si el proceso está en control, si no que adicionalmente ayudan a determinar si hay corrimientos en la media que puedan generar una señal futura, por cual se prefieren en lugar de las cartas por atributos que monitorean sólo si la variable está o no está dentro de un intervalo de especificación.
- Cuando se presenta una proporción de censura del 50% las tres cartas propuestas y la tipo Shewhart ignorando censura tienen una potencia similar en la detección de cambios en la media para cualquier tamaño de muestra, por lo que cualquiera de estas podría ser implementada para el caso de censura baja a moderada.

- Finalmente, Bajo la fase II del control estadístico, si se tienen datos censurados a intervalo y se desea conocer si la media del proceso está bajo control, se tendrá que la carta CEV- $\bar{X}$  es la más indicada, pues considera la censura presente en las observaciones y mayor probabilidad de detectar desviaciones en la media del proceso que incluso la carta Shewhart tradicional. Lo anterior deja por sentado el cumplimiento de los objetivos de esta tesis, pues se logró construir una carta de control eficiente para este tipo de censura, cuyo desempeño fue comparado y resultó mayor que en la carta  $\bar{X}$  Shewhart tradicional.

### 5.3. Trabajo futuro y recomendaciones

- Las cartas propuestas son para fase II de control, donde se asumen parámetros conocidos, y si estos no se conocen deben ser estimados basados en una muestra “grande” inicial, en la cual es necesario obtener observaciones exactas para poder tener información con la cual estimar, por lo que en fase I se debe medir la característica de interés en todas las unidades. También es necesario estudiar el efecto de la estimación de parámetros en las cartas propuestas y ahondar en el estudio del desempeño de las cartas frente a diversos tipos de corrimientos en la media, para así poder generalizar las conclusiones de los ejemplos presentados.
- Bajo el escenario de censura a intervalo en fase II de control, será de interés desarrollar también cartas de control para una distribución diferente a la normal en las observaciones u otros métodos de imputación, como el límite inferior del intervalo de censura. Adicionalmente, se puede estudiar si los resultados obtenidos en esta tesis se mantienen para tamaños de muestra diferentes a los considerados.
- Se puede pensar en implementar otras cartas de control, como las tipo CUSUM o EWMA para el monitoreo de la media, las cuales han demostrado tener mejor desempeño en detectar pequeños cambios sostenidos en la variable de interés, al menos cuando se trabaja con datos completos (no censurados). También proponer cartas de control para monitorear la desviación estándar del proceso bajo censura tipo I a intervalo en fase II, estudiar el desempeño y comparar su eficiencia con las cartas propuestas para el monitoreo de la media, determinando así qué carta usar para el monitoreo del proceso.
- Bajo censura a intervalo se puede recurrir al monitoreo del proceso mediante una carta por atributos, donde se estudie si la unidad cumple o no con las especificaciones, pero en este caso no se logra estudiar si la media del proceso ha presentado corrimientos y esté a punto de salirse de control estadístico. Por tanto, carta  $\bar{X}$  tipo Shewart se prefiere, pues permite detectar posibles cambios en la media durante el monitoreo del proceso.



# A Optimización usando el paquete “DEoptim” en R

En la librería `DEoptim` de R creada por D. Ardia, K. Mullen, B. Peterson and J. Ulrich, se encuentra la función del mismo nombre `DEoptim`, con la cual es posible la optimización global de una función objetivo usando algoritmos de evolución diferencial, cuya sintaxis es:

```
DEoptim(fn, lower, upper, control = DEoptim.control(), ..., fnMap=NULL)
```

Donde,

- `fn` es la función a ser optimizada (minimizada).
- `lower, upper` son dos vectores especifican límites inferior y superior de búsqueda de cada parámetro a optimizar.
- `control` consiste en una lista de parámetros dentro de `DEoptim.control()` que modifican la forma como es aplicado el algoritmo evolutivo.

Dentro de `DEoptim.control` se definen:

- `VTR` es el valor objetivo o a alcanzar, por defecto es `-inf`.
- `strategy` define la estrategia de evolución diferencial a usar en el procedimiento de optimización. Por defecto se tiene la estrategia 2, donde se toman son comparados el *i*-ésimo elemento y el mejor de la población anterior para continuar la evolución.
- `bs` si es igual a `FALSE` compara un miembro mutante contra uno de la generación anterior y deja el mejor.
- `NP` representa el número de miembros de la población. Por defecto es 50. Si se agranda hay mayor probabilidad de encontrar el mínimo pero toma mayor tiempo.
- `intermax` es el máximo de interacciones permitidas. Por defecto es 200.
- `CR` es la probabilidad de cruce entre miembros, cuyo valor está entre `[0,1]`. Por defecto es 0.5.
- `F` es el factor de ponderación diferencial, cuyo valor está entre `[0,2]`. Por defecto es 0.8.

- `trace` por defecto es `TRUE` y muestra los resultados en cada iteración.

Cabe notar que el resultado de `DEoptim` es una variable aleatoria, por lo que en cada ejecución se obtendrán valores diferentes, debido a lo cual se debe crear una semilla, por ejemplo `set.seed(1234)`. Con el fin de optimizar el logaritmo de la verosimilitud con censura a intervalo se define la siguiente función a ser optimizada:

```
f=function(x,datos,a,b){
mu=x[1]
sigma=x[2]
za=(a-mu)/sigma
zb=(b-mu)/sigma
ind=ifelse(datos>a & datos <b,0,1)
r=sum(ind)
n=length(datos)
l=r*log(sigma)-sum(log(dnorm(((datos-mu)/sigma)))*ind)-(n-r)*log(pnorm(zb)-pnorm(za))
1
}
```

Como `DEoptim` busca minimizar la función, el logaritmo de la verosimilitud debe ser ingresado negativo, para así lograr maximizarlo. Luego se aplica a la función,

```
library(DEoptim)
set.seed(1234)
DEoptim(f, lower, upper, DEoptim.control(trace=F,iter=1000),datos,a,b)$optim$bestmem
```

Donde, `datos` representa las observaciones luego de aplicar algún método de imputación para la censura, es decir son los  $w_i$ , y  $a, b$  son los límites inferior y superior de censura respectivamente, por medio de los cuales es determinada la proporción de censura. Para el cálculo los  $w_i$  iniciales en el caso de la carta CEV- $\bar{X}$  se puede recurrir al uso de la media y varianza muestral. Se debe usar `$optim$bestmem` al aplicar la función, con el fin de extraer el mejor conjunto de parámetros óptimos encontrados, en este caso los valores estimados para  $\mu$  y  $\sigma$ .

## B Optimización usando función “optim” en R

En R se encuentra disponible la función `optim`, la cual los valores de los parámetros de interés que minimizan una función objetivo. Su sintaxis es la siguiente:

```
optim(par, fn, method = c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B",  
                          "SANN", "Brent"), lower, upper, control = list())
```

Donde,

- `par` son valores iniciales de los parámetros a optimizar.
- `fn` la función a ser optimizada. Por defecto se busca el mínimo.
- `method` define el método a ser usado. Por defecto es el método expuesto por Nelder y Mead (1965), que utiliza únicamente valores de función y es robusto pero relativamente lento, aunque trabaja razonablemente bien para funciones no diferenciables. También está el método "L-BFGS-B" dado por Byrd et. al. (1995), donde a cada variable se le puede dar un límite inferior y / o límite superior. Entre otros.
- `lower`, `upper` son dos vectores que especifican límites inferior y superior de búsqueda de cada parámetro a optimizar para el método "L-BFGS-B".

Luego se aplica `optim` a la función objetivo:

```
optim(par, fn, lower, upper, method = "L-BFGS-B", datos, a, b,)$par
```

Donde `datos` representa las observaciones luego de aplicar algún método de imputación para la censura, es decir son los  $w_i$ , `a`, `b` son los límites inferior y superior de censura respectivamente. Se puede usar el método "L-BFGS-B", pues permite definir `lower` y `upper` para la búsqueda. La implementación de la función `optim` es más sencilla que `DEoptim`, pues requiere la definición de un menor número de criterios a la hora de optimizar. En el caso de tener una censura cercana al 100 %, la función es capaz de estimar siempre y cuando se inicialice la búsqueda con un valor para  $\sigma$  diferente de cero.

# Bibliografía

- [1] Ardia D., Boudt K., Carl P., Mullen K. M. and Peterson B. (2011). Differential Evolution with DEoptim. *The R Journal*, Vol. 3, pp.27-34
- [2] Brill R. V. y Bzik T. J. (2009). Control Charts for Skewed and Left-Censored Data that Approximately Control the Family Type I Error. *Quality Engineering*, Vol. 21, pp.446-462.
- [3] Crowder M., Kimber A., Smith R. y Sweeting T. (1994). *Statistical Analysis of Reliability Data*, Chapman y Hall, New York.
- [4] Das N. (2009). A Comparison Study of three Non-parametric Control Charts to Detect Shift in Location Parameters. *International Journal of Advanced Manufacturing Technology*, Vol. 41, pp. 799–807.
- [5] Feltz C. J. y Sturm G. W. (1994). Real-Time Empirical Bayes Manufacturing Process Monitoring for censored data. *Quality and Reliability Engineering International*, Vol. 10, pp.467-476.
- [6] Hsu C., Taylor M. G., Murray S. y Commenges D. (2006). Multiple Imputation for Interval Censored Data with Auxiliary Variables. *Statistics in Medicine*, Vol. 26, pp.769-781.
- [7] Hyndman R. J. y Fan Y. (1996). Sample Quantiles in Statistical Packages. *The American Statistician*, Vol. 50, pp.361-365.
- [8] Lin Feng-Tse. (2013). Control Parameter Settings in Differential Evolution for Global Optimization Problems *International Journal of Intelligent Technologies and Applied Statistics*, Vol. 6, pp.35-49.
- [9] Meeker W. y Escobar, L. (1998). *Statistical Methods for Reliability Data*, John Wiley y Sons, New York.
- [10] Pascal, F. y Li, S. (2011). Monitoring the Weibull Shape Parameter by Control Charts for the Sample Range of Type II Censored Data. *Quality and Reliability Engineering International*, Vol. 28, pp. 233-246.

- 
- [11] Ryan, T. P. (1989). *Statistical Methods for Reliability Improvement*, John Wiley y Sons, New York.
- [12] Sampford, M. R. y Taylor, J. (1959). Censored observations in Randomized Block experiments *Journal of Royal Statistical Society*, Vol. 21, pp.114-237.
- [13] Steiner, H. S. y Mackay, J. R. (2000). Monitoring Processes with Highly Censored data. *Journal of Quality Technology*, Vol. 32, pp.199-208.
- [14] Steiner, H. S. y Mackay, J. R. (2001). Monitoring Processes with Data Censored Owing to Competing Risks by Using Exponentially Weighted Moving Average Control Chart. *Journal of the Royal Statistical Society*, Vol. 50, pp. 293-302.
- [15] Storn, R. y Price, K. (1997). Differential Evolution. A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, Vol. 11, pp 341-359.
- [16] Taguchi G. (1986). *Introduction to Quality Engineering: Designing quality into products and processes*, Asian productivity organization.
- [17] Vargas, J. y Montaña, T. (2005). Carta de Control CEV X para Distribuciones Weibull con Datos Censurados. *Revista Colombiana de Estadística*, Vol. 28, pp. 125-139.
- [18] Woodall, W. H. (2000). Controversies and Contradictions in Statistical Process Control. *Journal of Quality Technology*, Vol. 32, pp. 341-350.
- [19] Teoh W. L., Khoo M. B. y Teh S.Y.(2013). Optimal Designs of the Median Run Length Based Double Sampling  $\bar{X}$  Chart for Minimizing the Average Sample Size. *Plos one Journal* Vol. 8, pp. 1-11.
- [20] Zhang, L. y Chen, G. (2004). EWMA Charts for Monitoring the Mean of Censored Weibull Lifetimes. *Journal of Quality Technology* Vol. 36, pp. 321-328.
- [21] Zhang, J. y Sanderson, A. (2009). *Adaptive Differential Evolution: A Robust Approach to Multimodal Problem Optimization*, chapter Theoretical Analysis of Differential Evolution.