

UNIVERSIDAD DE LOS ANDES
FACULTAD DE CIENCIAS ECONÓMICAS Y SOCIALES
INSTITUTO DE ESTADISTICA APLICADA Y COMPUTACION
DOCTORADO EN ESTADÍSTICA
MÉRIDA – VENEZUELA



www.bdigital.ula.ve

***“ÍNDICES DE CAPACIDAD DE PROCESOS PARA VARIABLES QUE
SE DISTRIBUYEN NO NORMAL MEDIANTE LA DISTRIBUCIÓN
PARETO GENERALIZADA”***

TUTOR: Dr José Simón Fermín

Autor: Martha Rosa Valdiviezo M.

TESIS PRESENTADA COMO REQUISITO PARCIAL PARA OPTAR AL GRADO DE
DOCTORA EN ESTADÍSTICA

Mérida, Julio de 2012

c.c Reconocimiento

AGRADECIMIENTO

A mi Dios por estar siempre iluminando mi camino.

A la Virgen del Carmen por llevarme de su mano, guiándome siempre.

A mi Madre por su apoyo y dedicación en todo momento.

A mi hija Kamila, cuyo amor infinito contribuyó enormemente en esta última fase de estudio.

A todos mis hermanos, por el apoyo brindado.

A mi tutor por su colaboración, sus sabios consejos, su tiempo, pero también gracias por la amistad.

A los Miembros del Jurado por las valiosas sugerencias realizadas al revisar este trabajo.

A los profesores del IEAC, quienes contribuyeron en mi formación académica.

A la Señora Lucy y Clemencia, gracias por la colaboración que me prestaron en todo momento.

www.bdigital.ula.ve

c.c Reconocimiento

INDICE GENERAL

	Página Nro.
AGRADECIMIENTO	2
INDICE GENERAL	3
INDICE DE TABLAS	7
INDICE DE FIGURAS	9
RESUMEN	10
CAPÍTULO 1.	
Introducción	11
Antecedentes	12
Problema de investigación	17
Justificación	18
Enfoques metodológicos	18
Contribuciones de la Investigación	21
Objetivos de la Investigación	
Objetivo general	22
Objetivos Específicos	22
Organización de la tesis	23
Aportes obtenidos de la Investigación	23
CAPÍTULO 2: Capacidad de procesos para variables de calidad distribuidas no normales	
2.1 Introducción	25
Índices de Capacidad de Procesos univariantes para datos no normales	25
2.3 Métodos para estimar ICP utilizando datos distribuidos no normales	28
2.3.1 Método percentiles de Clements	29
2.3.2 Método de percentiles de Burr	31
2.3.3 Método exacto	33
2.3.4 Método ICP basado en la función de distribución de probabilidad	34

2.3.5 Transformación de potencia de Box-Cox	35
2.3.6 Transformación de Johnson	36
2.3.6.1 Sistema de Transformación Log-normal (S_L)	36
2.3.6.2. Sistema no acotado (S_U)	37
2.3.6.3. Sistema acotado (S_R)	37
2.3.7. Varianza Ponderada	38
2.3.8 Índice de Capacidad de Procesos de Wright, (C_s)	39
2.3.9. Intervalos de Confianza para los estimados de ICP: C_p y C_{pk}	39
2.3.9.1 Estándar Bootstrap	40
2.3.9.2 Percentil Bootstrap	41
2.3.9.3 Percentil Sesgo-Corregido Bootstrap	41
CAPITULO 3: Ajuste de distribuciones con datos provenientes de variables de calidad no normales	
3.1 Introducción	42
3.2 Ajuste de distribuciones conocidas a datos provenientes de variables de calidad no normales	42
• Distribución Weibull	43
• Distribución Log-Normal	44
• Distribución Beta	44
• Distribución Pareto Generalizada	45
• Distribución Pareto Generalizada con 2 parámetros	47
3.3 Algoritmo genético.	49
3.3.1 Funcionamiento	50
3.3.2 Cuándo usar estos algoritmos	51
3.3.3 Operación de un algoritmo genético	52
CAPÍTULO 4: Análisis de Capacidad de Procesos Univariantes. Aplicaciones	
4.1 Introducción	54
4.2 Estimación de Capacidad de Procesos usando distribuciones no normales. Aplicación.	55
4.2.1 Estudio de simulación	57
4.2.2. Estimación y análisis de los ICP mediante diferentes métodos	58

4.2.2.1 Metodología de simulación.	58
Comparación de resultados	63
4.2.2.2 Caso Real	64
Intervalos bootstrap para ICP usando la DPG2	67
Métodos Bootstrap	67
a. Estándar Bootstrap: (EB)	67
b. Percentil Bootstrap: (PB)	68
c. Percentil Sesgo-Corregido Bootstrap: (PSCB)	68
• Estudio de simulación	69
• Simulación de datos	69
• Discusión de resultados	75
4.3 Índices de Capacidad estimados mediante la función de densidad de la DPG2	77
• Aplicación del modelo de la DPG	77
4.4 Índices de capacidad de procesos estimados mediante el método exacto	78
4.4.1. ICP estimados mediante el método exacto usando DPG2	79
4.4.2. Análisis de simulaciones	79
4.4.3. Resultados y conclusiones	86
CAPÍTULO 5: Índices de Capacidad de procesos basados en la DPG3. Aplicación.	
Introducción	88
5.2 Distribución generalizada de Pareto con tres y con cuatro parámetros	88
5.2.1 Distribución Pareto Generalizada con 3 parámetros	88
5.2.2 Distribución Pareto Generalizada con 4 parámetros	90
5.3. Estimación de parámetros para la DGP3 mediante el método de máxima Verosimilitud	93
5.3.1 Método de Máxima Verosimilitud para la DPG con 3 parámetros	93
5.3.2. Método de Momentos para la DPG con 3 parámetros	95
5.4. Desempeño del método de estimación de momentos de la distribución GP3	96
5.5. Algoritmo genético para estimar los parámetros de la distribución GP3	100

5.5.1 Sistema de ecuaciones no lineales escrito en R.	100
5.5.2. Algoritmo genético. Aplicación.	101
5.6. Métodos para estimar índices de capacidad de procesos usando la DPG3	103
5.6.1. Método ICP usando una función de distribución acumulada	103
5.6.2 Método ICP basado en la función de distribución de probabilidad	104
5.6.3 ÍCP basados en el método exacto considerando DPG3	105
5.7 Estimación de Índices de capacidad de procesos con una variable DPG3	105
5.7.1 Resultados de la estimación de ICP usando métodos no normales.	106
5.8. Índices de capacidad basados en el método exacto considerando la DPG4	109
5.9 Índices de Capacidad basados en la función densidad de la DPG4	109
CONCLUSIONES	111
RECOMENDACIONES Y FUTUROS TRABAJOS	114
BIBLIOGRAFÍA	115
ANEXOS	126

www.bdigital.ula.ve

INDICE DE TABLAS

	Página Nro.
Tabla 1.1: Proporción mínima de ítems no conformes (PNC)	16
Tabla 4.1: Resultados promedios para la Distribución Wei(1,0;1,2)	59
Tabla 4.2: Resultados promedios para la Distribución Be(14,4;13,3)	60
Tabla 4.3: Resultados promedios para la Distribución Log-normal(0;1)	60
Tabla 4.4: Resultados promedios para la Distribución PG2(-0.2;1.0)	61
Tabla 4.5: Resumen estadísticos para una serie de datos obtenidos en un proceso hidrológico con tiempo de duración de 1 hora de lluvia	66
Tabla 4.6: Valores de los índices de capacidad obtenidos con los métodos de percentil de Clements, de Burr y los tradicionales.	66
Tabla 4.7: Resumen estadístico de la muestra de observaciones provenientes de la DPG2(20,66; 0,44).	69
Tabla 4.8: Valores para diferentes C_{pu} , U_t y los índices de capacidad estimados	70
Tabla 4.9: Proporción de cubrimiento del 95% para \widehat{BC}_{pk}	71
Tabla 4.10: Proporción de cubrimiento del 95% para \widehat{CC}_{pk}	72
Tabla 4.11: Proporción de cubrimiento del 95% para \widehat{C}_{pk}	73
Tabla 4.12: Resultados de los ICP con una muestra simulada mediante el modelo PG2(n=25, α_n , 1)	80
Tabla 4.13: Resultados de los ICP con una muestra simulada mediante el modelo PG2(n=50, α_n , 1)	82
Tabla 4.14: Resultados de los ICP con una muestra simulada mediante el modelo PG2(n=100, α_n , 1)	84
Tabla 5.1: Casos poblacionales de la DPG3.	96
Tabla 5.2: Valores estimados de la DPG3 para cada conjunto de parámetros $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ (Cv=0,5 y Asim= 0,5)	97

Tabla 5.3: Sesgo para los parámetros: $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ ($Cv=0,5$ y $Asim= 0,5$)	98
Tabla 5.4: Varianza para los parámetros: $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ ($Cv=0,5$ y $Asim= 0,5$)	99
Tabla 5.5: Cuadrado medio del error para los parámetros: $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ ($Cv=0,5$ y $Asim= 0,5$)	99
Tabla 5.6: Valores estimados de la DPG3 para cada conjunto de parámetros $\alpha=-0,069$, $\beta=0,433$ y $\gamma=0,536$ ($Cv=0,5$ y $Asim= 2,5$)	99
Tabla 5.7: Sesgo para cada conjunto de parámetros $\alpha=-0,069$, $\beta=0,433$ y $\gamma=0,536$ ($Cv=0,5$ y $Asim= 2,5$)	100
Tabla 5.8: Resultados obtenidos mediante algoritmo genético y R.	103
Tabla 5.9: Resumen estadístico para la DPG3 con $n=100$ y el conjunto de parámetros: $\hat{\alpha}= 0,554$; $\hat{\beta} = 1,116$ y $\hat{\gamma} = 0,277$.	106
Tabla 5.10: Valores de los Cuantiles de la DPG3($0,554$; $1,116$; $0,277$)	106
Tabla 5.11: Valores de los ICP usando el modelo DPG3($0,554$; $1,116$; $0,277$), método de Percentil de Clements para diferentes valores de U_t	106
Tabla 5.12: Valores de los ICP usando el modelo DPG3($0,554$; $1,116$; $0,277$), método de Percentil de Burr para diferentes valores de U_t	107
Tabla 5.13: Valores de los ICP usando el modelo DPG3($0,554$; $1,116$; $0,277$), método de Percentil de Clements para diferentes valores de U_t	107
Tabla 5.14: Valores de los ICP usando el modelo DPG3($0,554$; $1,116$; $0,277$), método Exacto para diferentes valores de U_t	107
Tabla 5.15: Valores de los ICP usando el modelo $N(0,94; 0,49)$ para diferentes valores de U_t	108

INDICE DE FIGURAS

	Página Nro.
Figura 3.1: Función de densidad de probabilidad de la DPG con $\beta=0,5$; $\alpha=0,75$	48
Figura 3.2: Función de densidad de probabilidad de la DPG con $\beta=1,0$ y $\alpha=1,25$	49
Figura 4.1: Distribuciones Log-N(0,5;0,2)	55
Figura 4.2: Distribución Gamma(4,0; 0,5)	56
Figura 4.3: Distribución Weibull(1,0; 1,2)	56
Figura 4.4: Distribución PG(20,7; 0,04)	57
Figura 4.5: Diagrama de caja para los índices (CCp; BCp) y (CCpk; BCpk) respectivamente para una distribución Weibull(1,0;1,2)	61
Figura 4.6: Diagrama de caja para los índices (CCp; BCp) y (CCpk; BCpk) respectivamente para una distribución Beta(4,4;13,3)	62
Figura 4.7: Diagrama de caja para los índices (CCp; BCp) y (CCpk; BCpk) respectivamente para una distribución Log-normal(0;1)	62
Figura 4.8: Diagrama de caja para los índices (CCp; BCp) y (CCpk; BCpk) respectivamente para una distribución PG2(-0,2;1,0)	63
Figura 4.9: Histograma de la serie de datos con tiempo de 1 hora máxima de precipitación	66
Figura 4.10: Histograma de la muestra proveniente de la DPG2(20,66;0,44)	70
Figura 4.11: Cubrimiento de los intervalos Percentil Bootstrap	73
Figura 4.12: Cubrimiento de los intervalos Estándar Bootstrap	74
Figura 4.13: Cubrimiento de los intervalos Percentil Sesgo-Corregido Bootstrap	74
Figura 5.1: Función DPG3 el conjunto de parámetros $\alpha=0,554$; $\beta=1,116$; $\gamma=0,277$	98

RESUMEN

El presente trabajo de investigación, muestra la aplicación del modelo de la Distribución Pareto Generalizada en la estimación de índices de capacidad de procesos; se consideraron otras variables de calidad distribuidas Weibull, Log-normal y Beta para la estimación de los ICP (C_p y C_{pk}). Para evaluar su desempeño, se estimaron previamente los parámetros de la DPG mediante el método de máxima verosimilitud y el método de momentos. Se simularon muestras y finalmente se estimaron los ICP no normales mediante los métodos de Percentiles de Clements, percentiles de Burr, el método exacto y usando la función de probabilidad de la distribución Pareto Generalizada. La necesidad surge por las siguientes carencias:

- Literatura que involucre el estudio acerca de la precisión y estimación de los ICP cuando se toman en cuenta desviaciones moderadas y severas de la normalidad; es por ello que para solventar esto, se llevaron a cabo diferentes comparaciones de los métodos no normales, considerando datos simulados e incluso una aplicación real a la industria.
- La aplicación de la Distribución Pareto Generalizada con dos parámetros aplicados a algún tipo de proceso productivo.
- Un método de cálculo que permita la estimación de los tres parámetros de la Distribución Pareto Generalizada, ya que las expresiones matemáticas generadas durante el uso de este modelo resultaron ser, en su mayoría, muy áridas al optimizar y no se tiene bibliografía que permita determinar de manera directa los parámetros que conforman este modelo; los cuales, una vez conocidos, si es posible aplicar dicho modelo en el estudio de capacidad de procesos.

En el trabajo, aparte de los métodos de percentiles de Clements y percentiles de Burr, se aplican la metodología de estimación de ICP, presentada por Liu y Chen (2006) y el método exacto (Czarski (2008)), para estimarlos mediante la DPG con dos y tres parámetros. También se logró desarrollar un algoritmo genético que permitió estimar los parámetros de la DPG con tres parámetros con mejor precisión, en comparación con los métodos tradicionales.

Entre los resultados encontrados al estimar los ICP, se mencionan los siguientes:

En general, el método de percentiles de Burr es mejor ya que produce desviaciones menores con respecto a las obtenidas con el método de percentiles Clements, cuando el modelo usado es la DPG2; sin embargo, el método exacto genera resultados más próximos a los valores targets fijados en comparación a los métodos de percentiles de Clements y Burr. Por otro lado, al considerar las simulaciones para formar los intervalos bootstrap, se encontró que el ICP estimado con el método de percentiles de Burr, es mejor considerando el porcentaje de cubrimiento y el ancho promedio para todas las condiciones fijadas para C_{pu} . Más específicamente, para un valor target ($C_{pu}=0,50$), se observó que el intervalo percentil bootstrap es, en promedio, mejor estimador que el estándar bootstrap y el percentil sesgo-correctido. Para las otras condiciones sobre el C_{pu} ($C_{pu}=1,0$ y $1,5$), el intervalo percentil sesgo-correctido estima mejor los ICP mediante los métodos de Clements y Burr. Finalmente con el algoritmo genético se determinaron los valores de los parámetros de la DPG con 3 parámetros, y luego se simularon un conjunto de muestras. Posteriormente se estimaron los ICP generándose que el método de Clements arroja mejores resultados en comparación con los métodos de percentil de Burr y exacto.

CAPÍTULO 1

INTRODUCCION

Desde comienzos del siglo pasado, los índices de capacidad de procesos (ICP) han jugado un papel vital en el mejoramiento de la eficiencia operativa de los productos y procesos de fabricación lo que resulta en una reducción significativa de las posibles pérdidas en tales procesos (en productos, mano de obra, tiempo, materia prima), que ocurren debido al incumplimiento de las especificaciones del cliente. Aunque los índices de capacidad de procesos tales como C_p y C_{pk} están siendo ampliamente aplicados en la industria para la evaluación del desempeño de cualquier proceso existe una falta de entendimiento entre quienes utilizan estas medidas de capacidad y las suposiciones en las cuales se basan esencialmente la teoría estadística y la normalidad. Si los supuestos básicos son alterados, entonces la determinación de capacidad puede llevar a conclusiones erróneas (Deleryd, 1998).

La investigación llevada a cabo en este trabajo presenta métodos para evaluar los ICP cuando la distribución fundamental de la variable de calidad se distribuye no normalmente.

Es un hecho que el objetivo fundamental de toda medida de capacidad, es ayudar a decidir a los gerentes e ingenieros en aceptar o rechazar los resultados de un proceso basado en la conformidad a las especificaciones del consumidor. Esta investigación se ha enfocado por ese motivo, en presentar la eficacia de los métodos propuestos usando el criterio de la proporción de no conformidad (PNC), el cual es frecuentemente usado en la práctica para determinar la utilidad de los métodos de ICP.

En el resto de este capítulo se presentan los fundamentos de los índices de capacidad y se proporciona una revisión documental de esta investigación; los principales problemas y preguntas de investigación, los enfoques propuestos al problema de investigación, las contribuciones principales de esta investigación y un esbozo detallado del trabajo. De esta manera, se muestran tres publicaciones que son productos académicos de esta investigación.

ANTECEDENTES

El análisis de capacidad de procesos conjuntamente con el control estadístico de procesos y el diseño de experimentos, son métodos estadísticos que han sido usados por décadas con el propósito principal de reducir la variabilidad generada en muchos procesos industriales y productos (Albing, 2006). La necesidad de entender y controlar dichos procesos obtiene más y más relevancia debido a la creciente complejidad en los sistemas técnicos en la industria. Por otra parte, el uso de métodos estadísticos en la industria está también incrementándose por la introducción de conceptos de gerencia de calidad tal como el programa de seis sigmas, donde los métodos estadísticos que incluyen índices de capacidad de procesos, son parte importante (Hahn et al, 1999).

El análisis de procesos trata de determinar el desempeño de un proceso de manufactura, donde la información acerca de éste se utiliza para mejorar su capacidad. Además con el análisis de capacidad de procesos se puede determinar cuán bien el proceso se ejecutará en función de las especificaciones o requerimientos del producto. Sin embargo, antes de determinar la capacidad de un proceso es importante que el proceso esté estable y repetible. Esto es, que solamente estén presentes causas de variación (comunes) naturales. Cabe señalar que un análisis de capacidad de procesos se puede realizar incluso si el proceso es inestable. Pero el análisis dará una indicación de la capacidad en ese momento solamente y de aquí los resultados serán de uso limitado (Deleryd y Vannman, 1999).

Para comprobar si el proceso es estable, generalmente se aplica control estadístico de procesos (CEP). El propósito es detectar y eliminar, utilizando gráficas de control, las causas de variación asignables, esto para determinar si el proceso está en control estadístico y revela patrones sistemáticos en su salida. Una introducción de CEP se puede hallar en Montgomery (2005). En Valdiviezo y Fermín (2009) se describe de manera detallada una aplicación en la industria y se plantea un problema de monitoreo de un proceso univariante; mientras que en Fermín et al. (2009), se describe el mismo proceso pero tomando en cuenta el caso multivariante.

Cuando un proceso se encuentra estable, se pueden usar diferentes técnicas estadísticas con el fin de analizar posteriormente su capacidad (Montgomery, 2005). Por ejemplo, un histograma conjuntamente con estadísticos muestrales tales como promedios y desviaciones

estándar, arrojan información acerca del desempeño del proceso, mientras que la forma del histograma da información acerca de la distribución de la variable de calidad estudiada. Otra técnica simple es determinar la forma, centro y variabilidad de la distribución usando una gráfica de probabilidad normal. Cabe destacar que, las herramientas antes mencionadas solo proporcionan información aproximada solamente acerca de la capacidad del proceso. Sin embargo, el instrumento más frecuentemente usado cuando se ejecuta un análisis de capacidad son los denominados índices de capacidad de procesos (ICP). Un índice de capacidad de procesos es una medida sin unidad que cuantifica la relación entre el desempeño real del proceso y sus requerimientos especificados. En general, mientras el valor del índice es más alto, la cantidad de productos fuera de especificación resulta más bajo. Por otro lado, si el proceso no está produciendo un nivel aceptable de productos conformes, deberían iniciarse esfuerzos de mejoramiento. Estos esfuerzos se pueden basar en diseño de experimentos. Es decir, se puede por ejemplo, identificar la variable del proceso, que influencia la variable estudiada y hallar las direcciones para optimizar el resultado del proceso.

Los índices de capacidad de procesos (ICP), al igual que otros métodos estadísticos, se basan en supuestos fundamentales. Dos de estos supuestos más ampliamente usado en la industria son, la estabilidad y que la variable de calidad estudiada es independiente y distribuida normalmente (Albing, 2006). Para entender estos índices de capacidad convencionales, considere X_1, \dots, X_n como los valores presentes de una cierta variable de calidad lo cual corresponde a n ítems seleccionados aleatoriamente de un proceso de producción y supóngase que tal variable se sitúan entre dos valores denominados límites de especificación inferior (LEI) y superior (LES) para cumplir con las normas de ingeniería establecidas por la empresa.

Los ítems que caen fuera de estos límites son serán considerados no conformes. Los casos especiales donde solo se requiere un límite de especificación se obtienen cuando $LEI \rightarrow -\infty$ o $LES \rightarrow \infty$.

La capacidad de procesos se diseña básicamente para monitorear la proporción de ítems no conformes y al mismo tiempo evitar la salida excesiva de éstos. Ésto se realiza en periodos

de clasificación específicos, se asumen que no existen errores de mediciones y dichas medidas se toman para valores que son independientes e idénticamente distribuidos.

El análisis de capacidad tradicional es aplicable para evaluar los índices de capacidad y relaciona la variabilidad permisible del proceso (LES-LEI) con su variabilidad natural, por costumbre, tomada como 6σ (Bernardo, 1996). Bajo estos supuestos el índice más frecuentemente usado en la industria es el C_p , presentado por Juran (1974),

$$C_p = \frac{LES - LEI}{6\sigma} \quad (1.1)$$

Es decir, C_p es la razón de la variabilidad permisible a la variabilidad natural. En particular si el X_i esta normalmente distribuida $N(\mu, \sigma^2)$, y el proceso es centrado en la media nominal por,

$$m = \frac{LES + LEI}{2} \quad (1.2)$$

es decir, $E[X]=\mu=m$, entonces un proceso capaz, es un proceso para el cual $C_p \geq 1$, con a lo más 0,27 % de ítems no conformes, es decir, 2700 ítems no conformes por millón producido en un proceso de producción.

El índice de capacidad de procesos dado por la ecuación (1.1), solamente considera la variabilidad del proceso. Naturalmente que es posible tener cualquier proporción de ítems fuera de los límites de especificaciones con solo relocalizar la media del proceso, es por esto que el C_p solamente cuantifica el desempeño potencial de dicho proceso cuando este es centrado en el punto medio de los límites de especificaciones. Si el proceso no está centrado, pero es asimétrico a alguno de los dos lados superior o inferior de los límites de especificación, entonces se aplican los índices de capacidad de un lado (superior e inferior respectivamente). Estos índices de capacidad se presentan en las ecuaciones (1.3) y (1.4).

$$C_{pu} (C_p \text{ superior}) = \frac{LES - \mu}{3\sigma} \quad (1.3)$$

$$C_{pl} (C_p \text{ inferior}) = \frac{\mu - LEI}{3\sigma} \quad (1.4)$$

donde μ y σ son la media y desviación estándar del proceso en control, respectivamente.

Si el proceso no está centrado en el punto medio de las especificaciones de ingeniería (LEI, LES), entonces su desempeño es medido tradicionalmente por el C_{pk} . Este índice de capacidad de procesos fuera del centro (ecuación 1.5), fue definido por Kane (1986) y es el valor mínimo de los índices de capacidad superior o inferior,

$$C_{pk} = \min \left\{ \frac{LES - \mu}{3\sigma}, \frac{\mu - LEI}{3\sigma} \right\}, \quad (1.5)$$

el cual es una distancia normalizada entre la media del proceso y su límite de especificación más cercano.

De la distancia anterior es evidente que el índice de capacidad C_p , mide el rango permisible de medidas relacionadas al rango actual de medidas y el C_{pk} mide la distancia entre el valor esperado y el límite de especificación más cercano relacionado a la mitad del rango actual de medidas si la variable de calidad está distribuida normalmente y si el proceso está bien centrado, es decir, la media del proceso está localizada en el punto medio del intervalo de especificación de dos lados; esto es, $\mu = m = \frac{LES + LEI}{2}$. Un $C_p \geq 1$, implica que el número

de valores de la variable de calidad estudiada fuera de los límites de especificación será pequeño (Pearn y Kotz, 1994) y obviamente la proporción de no conformidad a las especificaciones del consumidor serán mínimas. En efecto, la razón de no conformidades es la principal interpretación del índice de capacidad de procesos C_p .

En este trabajo, la eficacia de los métodos propuestos serán presentados usando el criterio de no conformidad. Por eso es necesario entender la definición y las relaciones con el C_p . Es interesante notar que el C_p en su cálculo tiene un enlace directo con las proporciones de ítems que caen fuera de los límites de especificaciones (Telmoudi, 2005). Se considera una variable de calidad bajo el supuesto de normalidad, y la proporción de ítems no conformes (PNC), se indican más abajo.

$$P[X > LES] + P[X < LEI] = P\left[\frac{X - \mu}{\sigma} > \frac{LES - \mu}{\sigma}\right] + P\left[\frac{X - \mu}{\sigma} < \frac{LEI - \mu}{\sigma}\right]$$

$$P[\text{PNC}] = 1 - \Phi\left[\frac{\text{LES} - \mu}{\sigma}\right] + \Phi\left[\frac{\text{LEI} - \mu}{\sigma}\right]$$

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \frac{e^{-u^2}}{2} du \quad (1.6)$$

Si μ , es sustituido por el punto medio de la variabilidad de la especificación $\left(\frac{\text{LES} + \text{LEI}}{2}\right)$ como se da en la ecuación (1.2), entonces la proporción de no conformes (PNC) se puede expresar como.

$$\text{PNC} = 2\Phi[-3C_p] \quad (1.7)$$

donde, $\Phi(x)$ es la función de distribución acumulada de la Unidad Gaussiana.

Es importante notar que el índice C_p , depende de la desviación estándar (σ) del proceso en control. En la tabla 1.1, se presenta la proporción de no conformidad para algunos valores dados de C_p (Kotz y Johnson, 1993). Se ha establecido el hecho de que, si el índice C_p excede 1,33 implica que los ítems caen fuera, los límites requeridos son más pequeños en números (63 partes solo caen fuera de las especificaciones por millón de partes producidas) y se considera que el proceso es capaz lo suficiente para llenar los requerimientos del consumidor. Contrario a esto, si el índice es más pequeño que 1,33 pero mayor que 1, se recomienda examinar el proceso a medida que el número de PNC se incrementa. Para valores del índice menor que 1, el proceso se considera incapaz.

Tabla 1.1. Proporción mínima de ítems no conformes (PNC)

C_p	2,0	$\frac{5}{3}=1,67$	$\frac{4}{3}=1,33$	1,00	$\frac{2}{3}=0,667$	$\frac{1}{3}=0,33$
PNC	0,002 ppm	0,57 ppm	63 ppm	2700 ppm	45500 ppm	317300 ppm

Ppm: denota partes por millón

Es evidente, de la discusión anterior, que la base intuitiva de los índices de capacidad de procesos C_p y C_{pk} depende del supuesto implícito de normalidad de la salida del proceso.

PROBLEMA DE INVESTIGACION

Es un hecho establecido que la calidad siempre ha sido parte integral de todos los negocios y servicios. En el ambiente competitivo de hoy en día, el paradigma del mejoramiento de la calidad ha cambiado, de ser una meta departamental a ser un objetivo “paraguas” para una organización. Muchas empresas incrementan los niveles de calidad de sus productos manufacturados más rápido, reduciendo la proporción de ítems no conformes dentro de las especificaciones, adaptándose a los requerimientos del consumidor para guiar y sostener su posición de mercado en esta era siempre cambiante.

Desde el advenimiento del control estadístico de procesos, las medidas de capacidad de procesos C_p y C_{pk} han jugado un rol importante en el mejoramiento de la calidad dentro de la industria. Sin embargo, como se discutió anteriormente, los índices de capacidad de procesos dependen ampliamente del supuesto que el proceso debe estar bajo control y estable; además el resultado de calidad del proceso debe ser independiente y distribuido normalmente. En el mundo real, los datos del proceso no siempre siguen una distribución normal. Numerosos estadísticos e ingenieros de calidad han llevado a cabo investigaciones sobre índices de capacidad de procesos normales; entre ellos, Zimmer y Burr (1963), Burr (1967), Clements (1989), Pyzdek (1992), Kotz y Johnson (1993), Boyles (1994), Wright (1995), Castagliola y Castellanos (2005), Somerville y Montgomery (1997), Chou y Cheng (1997), Kotz y Lovelace (1998), Deleryd (1998), Wu et al. (1998), Tang y Than (1999), Kotz y Johnson (2002), Liao et al. (2002), Liu y Chen (2006); pero el principal objetivo de sus investigaciones se concentró en procesos con una sola variable de calidad.

Para investigar las posibles soluciones a los ICP esta investigación intenta hallar las respuestas a las siguientes preguntas:

- ¿Cuáles son los métodos más utilizados para calcular la capacidad de procesos cuando se consideran datos provenientes de procesos no normales con solo una variable de calidad? ¿Es tomada en cuenta la necesidad de ejecutar una comparación respecto a su aplicabilidad a los problemas del mundo real?
- ¿Puede una distribución no normal como la distribución Pareto Generalizada ser utilizada con datos reales y ser capaz de proporcionar los “mejores” estimados de

índices de capacidad de proceso no-normales con datos univariantes? ¿Se puede aplicar al caso multivariante?

Las preguntas antes planteadas fueron tomadas como base para el desarrollo de este trabajo. Por otro lado, se aplicaron herramientas utilizadas dentro de la teoría estadística y el análisis numérico, lográndose la construcción de programas computacionales que permiten al usuario decidir cuál distribución no normal se puede aplicar en el campo de control estadístico de procesos utilizando ICP estimados.

JUSTIFICACION

En la revisión efectuada durante la investigación se encontraron diversos métodos que permiten el análisis de datos distribuidos Beta, Weibull, Gamma considerados para estimar índices de capacidad de procesos. No obstante, en este trabajo se estimaron los índices de capacidad de procesos utilizando una distribución no normal como lo es la distribución de Pareto Generalizada con 2 y 3 parámetros usando datos reales y simulados. Los resultados logrados aquí se compararon con otros resultados alcanzados usando otras distribuciones no normales.

Hasta ahora no se han encontrado resultados que involucren la estimación de ICP mediante la utilización del modelo de la DPG dentro del área de control estadístico de procesos, por lo que se piensa que es de importancia su desarrollo y aplicación dentro del área de Control de Calidad y alternativamente se presenta conjuntamente con otras distribuciones no normales como las distribuciones Beta, Log-normal, Weibull o Gamma las cuales son de gran aplicación en dicha área. Adicionalmente, se logró estudiar el desempeño de un proceso de calidad con la capacidad de decidir, de acuerdo a los programas computacionales desarrollados y presentados en los anexos de este trabajo, cuál distribución no normal es la adecuada, bajo diferentes circunstancias, para la evaluación del desempeño de un proceso industrial y en tal sentido esto enriquece esta área de investigación.

ENFOQUE METODOLOGICO PROPUESTO

Como se mencionó anteriormente, cuando los datos son no normales, la medición de capacidad de procesos usando métodos convencionales puede guiar a conclusiones erróneas. Diferentes índices de capacidad han sido propuestos para tratar con el tema de no

normalidad. Aunque estos métodos son aplicados en las industrias, no existe suficiente literatura para determinar la precisión del desempeño de un proceso considerando desviaciones moderadas y severas de normalidad. En función a esto, en este trabajo se evalúa el desempeño de los métodos de estimación de capacidad existentes para datos con variable de calidad univariantes no normales, es decir, método de percentil de Clements, método de percentil de Burr, el método exacto y el método de Box-Cox.

Se conduce un estudio de simulación usando distribuciones no normales conocidas, Weibull, Beta y log-normal con la intención de comparar el desempeño de los ICP.

Por otro lado, se estimaron los ICP: C_p y C_{pk} , utilizando el modelo de la distribución Pareto Generalizada con 2 parámetros (DPG2) en el área de control de calidad. A juicio de los autores, esta distribución no se ha utilizado todavía en el área de procesos productivos. Sin embargo, en este trabajo se utilizó dicho modelo y se logró la estimación de los ICP univariantes usando datos reales y simulaciones. Además, se utilizó el modelo Pareto Generalizado con 3 parámetros, para estimar los ICP univariantes. Previo a esto, se construyó un conjunto de programas computacionales que permiten estimar los parámetros de la DPG ya sea con 2 o 3 parámetros. Si bien es cierto, existe la teoría que permite determinar de manera directa, mediante fórmula, los parámetros de una DPG con 2 parámetros; para el caso de la DPG con 3 parámetros no ocurre así, es por ello que en este trabajo se presenta un programa usando un algoritmo genético, que permite estimar los parámetros de dicha distribución. Se utilizó la herramienta de algoritmo genético usando MatLab* por la limitación de otros métodos de análisis numéricos ante la resolución de las ecuaciones generadas con el método de momentos utilizado para la estimación de los parámetros de la DPG con 3 parámetros. Finalmente, se deja propuesto el uso del modelo de la DPG con 4 parámetros para estimar ICP y determinar su aplicabilidad en esta área.

Una vez obtenidos los parámetros de la distribución, el modelo queda definido, listo para simular las muestras de diferentes tamaños; se utilizan las fórmulas de los ICP con distribuciones no normales y se establece la comparación de los resultados usando los métodos de percentiles de Clements, percentiles de Burr y método exacto. Para ello, se construyeron diversos programas computacionales usando S-plus* y R® tanto en la

simulación de los datos como en la estimación de los parámetros de distintas distribuciones de probabilidad. Así mismo, se desarrollaron otros programas computacionales (ver anexos), que permiten estimar los ICP usando diferentes distribuciones no normales, para decidir cuál método logra la mejor estimación del ICP en diferentes situaciones de desviaciones de normalidad.

Otro método al cual se hace referencia en este trabajo es la relación existente entre el índice estimado bajo un enfoque bayesiano propuesto por Bernardo e Irony (1996) y el estimado bajo el método tradicional. Como ya se mencionó anteriormente, se asume inicialmente que el proceso es estable, de tal manera que las observaciones disponibles puedan ser tomadas como una muestra aleatoria proveniente de alguna distribución fundamental. A fin de verificar tal supuesto, se sugiere ver Bernardo y Smith (1994). Este índice de capacidad Bayesiano denotado por C_b , está directamente relacionado con los índices tradicionales (C_p y C_{pk}). Singpurwalla (1998), lo introduce de manera sistemática tomando en cuenta una transformación monótona y genera la siguiente relación.

$$\frac{1}{w} \Phi^{-1} \left(\Pr (x \in A | \underline{x}_i(n)) \right) \geq \frac{1}{w} \Phi^{-1} \left(\frac{b}{a} \right) \quad (1.8)$$

donde, Φ es la función de distribución normal estándar y w es un valor constante positivo. Además, se plantea la posibilidad de que Φ sea una transformación de la forma: $\log \left(\frac{v}{1-v} \right)$

Cuando $w=3$, la expresión (1.8) define el Índice de Capacidad Bayesiano (C_b), planteado por Bernardo e Irony (1996). Ellos plantean este índice de manera teórica y describen la relación entre C_b , C_p y C_{pk} .

En Valdiviezo y Fermín (2010), se tomó en cuenta este índice bayesiano y los tradicionales usando una aplicación real a la industria de alimento y una aplicación con datos simulados, estableciéndose así una comparación entre los valores resultantes de C_p , C_{pk} y C_b . La empresa facilitadora de los datos para esta aplicación fue la Industria Procesadora de Harina de Maíz (REMAVENCA Sucursal Cumaná). Se tomó en cuenta las apreciaciones de los expertos de calidad

de dicha empresa quienes facilitaron sus opiniones acerca de algunas variables que posiblemente afectaban el proceso de Laminación; sugirieron a la “humedad del endospermo” como la posible causa. Se verificó que sigue una distribución ligeramente normal, no centrada y además se observó que el 72% de los ítems producidos estaban dentro de los límites de especificaciones establecidos. Así la proporción esperada de ítems no conformes superó el 0,27% lo cual determinó finalmente que el proceso no es capaz y necesita revisión. Esto es, la variable no generó valores acordes a los estándares de calidad, pues los índices obtenidos, están por debajo del valor estipulado, lo que induce a pensar que el proceso se deba revisar.

Además de lo antes planteado, ¿qué ocurre cuando el método para estimar los ICP no es el tradicional? Esto se responde también en Valdiviezo Fermín (2010), donde se consideró la estimación de ICP usando la teoría Bayesiana. Allí se compararon resultados y se observó que el índice de capacidad estimado bayesiano, $\hat{C}_b(D) = 0,8099$ es mayor que el índice de capacidad estimado tradicional $\hat{C}_p = 0,46$, pero ligeramente más pequeño que el índice $\hat{C}_{pk} = 0,811$, en todos los casos considerados, dichos índices están por debajo de los estándares de calidad, lo que produce mayor porcentaje de ítems no conformes y en consecuencia se tiene un proceso de laminación no capaz.

Si bien es cierto que una sola variable no define el adecuado desempeño de un proceso, se sugiere considerar otras variables de calidad.

CONTRIBUCIONES DE LA INVESTIGACIÓN

Esta investigación se basó fundamentalmente en determinar, cómo afecta la condición de no normalidad de los datos en la estimación de los ICP, tomando en cuenta diferentes distribuciones asimétricas pero principalmente la distribución Pareto Generalizada. No obstante, se generan otras contribuciones logradas en el desarrollo del trabajo, como son las que se presentan a continuación:

1. La revisión bibliográfica de los métodos existentes para la estimación de índices de capacidad de procesos tanto normales como no normales.

2. En lugar de aplicar el método de momentos tradicional citado en la investigación, se utilizó el método de máxima verosimilitud para determinar la precisión de la estimación de ICP basado en el método de percentil de Burr usando la DPG con 2 parámetros.
3. Se propone un algoritmo genético para estimar los parámetros de la distribución Pareto Generalizada con 3 parámetros.
4. Otra contribución es el planteamiento de la función de probabilidad de la Distribución Pareto Generalizada (DPG) con dos parámetros en la estimación de ICP, así como la aplicación del método exacto como herramienta para la estimación de dichos índices considerando la distribución Pareto Generalizada con dos y tres parámetros. El enfoque propuesto es en contraste al enfoque adoptado en la literatura de investigación, es decir, usar el “mejor” ajuste de la función de densidad de probabilidad a los datos no normales para la estimación de ICP.

OBJETIVOS DE LA INVESTIGACIÓN

Objetivo General

Comparar métodos de estimación de ICP no normales utilizando estudios de simulación, tomando en consideración diversas variables de calidad distribuidas no normalmente.

Objetivos Específicos

- 1) Establecer las bases conceptuales referentes a los ICP univariantes para variables de calidad distribuidas tanto normal como no normal.
- 2) Establecer las bases conceptuales referentes a diferentes distribuciones no normales en particular la distribución Pareto Generalizada con 2 y 3 parámetros.
- 3) Estimar los parámetros de la distribución Pareto Generalizada mediante los métodos de momentos y máxima verosimilitud.
- 4) Comparar los métodos de ICP no normales, percentiles de Clements, percentiles de Burr, método exacto y función de densidad de probabilidad, realizando las pruebas con datos reales y simulaciones.

ORGANIZACIÓN DE LA TESIS

Los subsiguientes capítulos se han organizado de la siguiente manera:

El capítulo 2 cubre una revisión documental y principios básicos de medidas de capacidad de procesos para variables de calidad normales y no normales. También, se presentan métodos para estimar los ICP para variables de calidad univariantes no normales.

En el capítulo 3, se presenta una descripción de distintas distribuciones no normales, así como de la Distribución Pareto Generalizada, los métodos de estimación de sus parámetros y aplicaciones. Por otro lado, se presenta una revisión documental de los que es un algoritmo genético y su funcionamiento; esta herramienta es utilizada para la estimación de los parámetros necesarios de la DPG con 3 parámetros.

En el capítulo 4, se lleva a cabo la comparación de métodos de estimación de ICP. Este capítulo cubre métodos de estimación de ICP usando el modelo de la DPG con 2 parámetros. Se comparan los resultados usando datos reales y simulados.

En el capítulo 5. Se desarrolla una metodología para la estimación de los 3 parámetros de la DPG, usando el método de momentos. Una vez conocido el modelo de la distribución, se simulan las muestras necesarias para estimar los ICP. Para la estimación de los parámetros se usó un algoritmo genético.

Por otro lado, se comparan los resultados de los ICP estimados mediante los métodos de percentil de Clements, de Burr, la metodología usada por Liu y Chen (2006) y el método exacto. Finalmente, se incluye conclusiones y recomendaciones producto de la investigación realizada.

APORTES OBTENIDOS DE ESTA INVESTIGACIÓN

Como resultado de la investigación realizada se logró la publicación de los siguientes trabajos en diferentes revistas.

Valdiviezo, M. y Fermín, J. (2009). Evaluación del Proceso de Laminación del Endospermo de maíz mediante Índices de Capacidad de Procesos. Revista Industrial Data. Vol.12(1). No 1: 47-54. En este trabajo se hizo la estimación de los ICP usando el enfoque tradicional y el bayesiano, siguiendo la metodología propuesta por Bernardo e Irony(1996).

Se utilizaron muestras simuladas y una aplicación en un proceso productivo, el cual se lleva a cabo en una industria productora de alimentos.

Valdiviezo, M. y Fermín, J. (2010). Estudio comparativo de índices de capacidad de procesos con variables distribuidas no normales. *Revista Industrial Data*. Vol. 13(2). Pp: 56-65. En este trabajo se evalúa el desempeño de los ICP en situaciones no normales, usando los métodos de Clements y Burr. Entre las distribuciones consideradas están: Beta, Log-normal, Weibull y la Pareto Generalizada con dos parámetros; se destaca la aplicabilidad de esta última en una aplicación con datos reales y simulados.

Valdiviezo, M. y Fermín, J. (2010). Estimación de índices de capacidad de procesos usando la Distribución Generalizada de Pareto. *Revista de Ingeniería Industrial*. Universidad Bio-Bio. Vol 1(2). Pp: 93-106. En este trabajo se estudia el desempeño y confiabilidad del método bootstrap para estimar intervalos de confianza, como herramienta en la evaluación de los ICP usando el modelo de la DPG con 2 parámetros.

www.bdigital.ula.ve

c.c Reconocimiento

CAPÍTULO 2

CAPACIDAD DE PROCESOS PARA VARIABLES DE CALIDAD DISTRIBUIDAS NO NORMALES

2.1 INTRODUCCION

Los índices de capacidad de procesos son variables aleatorias y como es bien sabido, las variables aleatorias están siempre asociadas a una distribución de probabilidad. Esta distribución también proporciona una descripción del valor esperado y varianza de dichos índices (Kotz y Lovelace, 1998). Conservando esta noción en mente, es importante entender la teoría de probabilidad y los métodos estadísticos a fin de entender completamente la naturaleza y conducta de los índices de capacidad de procesos presentados en la literatura de investigación.

Este capítulo se enfoca en aquellos métodos y técnicas comúnmente usadas para estimar capacidad de procesos cuando las variables de calidad se distribuyen no normalmente.

2.2 ÍNDICES DE CAPACIDAD DE PROCESOS UNIVARIANTES PARA DATOS NO NORMALES

La base de los índices de capacidad de procesos tradicionales C_p y C_{pk} depende de un supuesto implícito de que las medidas de las variables de calidad son independientes y distribuidas normalmente. Sin embargo, este supuesto básico no se cumple a cabalidad en la práctica. Muchos procesos físicos producen variables de calidad no normales y los investigadores necesitan verificar dichos supuestos básicos antes de desplegar cualquier método convencional para estimar índices de capacidad de procesos y determinar la capacidad de dicho proceso.

Kane (1986), puso su atención en los problemas que pueden surgir con datos no normales, luego Gunter (1989), evidenció esto aún más. Para evitar estos problemas han surgido varios enfoques. En la sección 2.3 se discuten dos enfoques comunes, la técnica de estimación de variables no normales y la de transformación. Adicionalmente, en el capítulo 3 se engloba el tema de no normalidad mediante el ajuste de distribuciones no normales conocidas a datos provenientes de variables de calidad no normales. Para una minuciosa

discusión de diferentes métodos utilizados en el manejo de resultados de procesos distribuidos no normalmente se sugiere ver Kotz y Johnson (1993), Kotz y Lovelace (1998) y Kotz y Johnson (2002).

Uno de los primeros índices para datos con distribución no normal fue sugerido por Clements (1989). Él usó la técnica de estimación de cuantiles no normales y propuso que 6σ y μ en el C_p y C_{pk} fueran remplazadas con $q_{0,99875}$, $q_{0,00135}$ y $q_{0,5}$ respectivamente, donde los q_α para valores de α especificados, representan los cuantiles para una distribución en la familia de Pearson. Si la distribución de la variable de calidad se distribuye normalmente, entonces $q_{0,99875} - q_{0,00135} = 6\sigma$. Se discute este método en la sección 2.3.1, Pearn y Kotz, (1994), extendieron el método de Clements para desarrollar algunos índices nuevos, los cuales pueden ser utilizados para estimar los ICP no normales.

Otro enfoque para tratar con datos no normales, es transformar los datos originales no normales a normales o al menos aproximadamente normal (en la sección 2.3 se darán detalles de técnicas de transformación). Gunter (1989), sugirió la aplicación del enfoque de la transformación de datos para ejecutar cálculos de índices de capacidad de procesos, cuando los datos son no normales. El cálculo de C_{pk} para datos no normales son discutidos en Rivera, Hubele y Laurence (1995). Posteriormente, Polansky, Chou y Mason (1998, 1999) propusieron un método para determinar la capacidad de procesos usando datos provenientes de una distribución normal truncada, donde la transformación de Johnson (Johnson, 1949) fue utilizada para transformar los datos del proceso no normal en normal. Sin embargo, no se puede estar seguro que la capacidad de la distribución transformada reflejará la capacidad de la distribución verdadera en una manera correcta (Gunter, 1989). Adicionalmente, Kotz y Lovelace (1998), señalaron que un profesional en el área de calidad puede trabajar incómodo con datos transformados debido a las dificultades en trasladar los resultados del cálculo transformados a la escala original. Otra desventaja desde el punto de vista de un profesional en el área de calidad, es que las transformaciones no relacionan lo suficientemente a las especificaciones originales de acuerdo con Kotz y Johnson (2002).

Además de los enfoques de percentiles y transformaciones, otros métodos han sido propuestos en la literatura. Para el caso con distribuciones asimétricas y límites de especificación de dos lados, Wu et al. (1999), introdujeron un nuevo índice de capacidad de procesos basados en un método de varianza ponderada. La principal idea de este método es dirigido a una distribución asimétrica en dos distribuciones normales con sus medias, para crear dos nuevas distribuciones, las cuales tienen la misma media pero diferentes desviaciones estándares.

Chang et al. (2002), propusieron un método diferente de construir simples índices de capacidad de procesos para poblaciones asimétricas, basados en un método de desviación estándar ponderada. Algunas propiedades para los índices propuestos fueron también investigados por Wu et al. (1999) y Chang et al. (2002) y los estimadores fueron comparados con otros métodos para datos no normales.

Diversos autores han realizado estudios comparativos entre diferentes métodos para manipular datos de procesos no normales. Heuvel y Ion (2003), compararon índices para distribuciones asimétricas propuestas por Munechika (1986), Bai y Choi (1997) para un número de distribuciones correspondientes a C_{pk} . Una conclusión de su estudio es que para muchas situaciones prácticas, el valor verdadero de C_{pk} , cae entre los valores de los índices presentados por Munechika, Bai y Choi. Usando simulaciones de Montecarlo, We et al. (2001) compararon índices tradicionales para métodos de Clements, método de Johnson-Kotz-Pearn (1992) y el método de varianza ponderada (Wu et al. 1999) para la familia de distribuciones de Johnson. Encontraron que para casos acotados asimétricos ninguno de estos tres métodos se ejecuta bien al estimar el valor nominal. Para casos log-normal, el método de varianza ponderada subestima los valores nominales, mientras que el método de Johnson-Kotz-Pearn sobreestima consistentemente los valores nominales. El método de Clements ni sobreestima, ni subestima los resultados en una base consistente.

Capacidad de procesos para límites de especificación de un lado cuando los datos son no normales

Los índices de capacidad de procesos para límites de especificación de un solo lado y una variable de calidad distribuida no normalmente no se ha discutido mucho en la literatura ya que esto no es una situación muy común en la industria. Sin embargo, Clements (1989), estudió tanto el C_p como el C_{pk} tomando en cuenta el caso cuando el límite de especificación está en un solo lado; reemplazó en la ecuación 1.3 y 1.4 el valor de μ por la mediana, el denominador por la diferencia entre los puntos porcentajes inferior (Percentil 0,135) y superior (percentil 99,865) de una distribución X . Sarkar y Pal (1997), consideraron una distribución de valor extremo para el caso C_{pu} , el cual representa un límite de especificación. Además Tang et al. (1999) estudiaron estimadores de C_{pu} para un número de métodos diseñados para manipular datos de procesos no normales, en particular, cuando las distribuciones son Weibull y log-normal. Esto fue hecho con simulaciones Montecarlo. Encontraron además que los métodos que involucran transformaciones, proporcionan estimados de C_{pu} más cercanos al valor nominal, comparado a métodos de no transformaciones como el método de varianza ponderada discutido por Choi y Bai (1997). Sin embargo, aun cuando un método ejecute bien para una distribución particular, ese método puede dar resultados erróneos para otra distribución con conducta de cola diferente. Es un hecho que, el efecto del área de la cola puede ser bastante dramática (Vannman y Albing, 2007).

Ding (2004), introdujo un índice de capacidad, basado en el rango efectivo y usando los primeros cuatro momentos con datos de un proceso no normal. Además de eso, introdujo el caso cuando los datos son asimétricos positivos univariantes y propuso un índice para este caso. Sin embargo, el índice propuesto no contiene valor objetivo (target) y adicional a esto, no se han presentados procedimientos de decisión o pruebas al respecto.

2.3 METODOS PARA ESTIMAR ICP UTILIZANDO DATOS DISTRIBUIDOS NO NORMALES

En esta sección se proporciona información acerca de algunos de los métodos comúnmente utilizados en la industria para la estimación de capacidad de procesos no normales como son el método de percentiles de Clements, el método de transformación de potencia de

Box-Cox, el método de percentil de Burr, el método exacto, la transformación de Johnson, la varianza ponderada y el índice de Wright.

2.3.1. Método de Percentiles de Clements

El sistema de curvas de Pearson está definido como el conjunto de funciones $f(x)$ que satisfacen la siguiente ecuación diferencial.

$$\frac{df(x)}{dx} = \frac{(x-a)f(x)}{b_0 + b_1x + b_2x^2} \quad (2.1)$$

donde a , b_0 , b_1 y b_2 son cuatro parámetros reales. Se puede mostrar que entre las soluciones para (2.1), se pueden hallar diferentes clases de curvas, como por ejemplo, distribuciones acampanadas, en forma de J y forma de U. Para detalles adicionales, se sugiere ver Elderton y Johnson (1969) y Kendal y Stuart (1969). Clements (1989), propuso un método para calcular índices de capacidad de procesos C_p y C_{pk} basados en distribuciones pearsonianas no normales. Este método proporciona una manera para manejar el tópico de no normalidad para la estimación de capacidad de procesos y usa percentiles no normales para modificar los índices de capacidad tradicionales. La principal ventaja de este enfoque es que dicho método no requiere el ajuste de distribuciones complicadas y su uso es simple. El método de Clements goza de mucha popularidad entre quienes practican la profesión de calidad en la industria. Este método usa las curvas de Pearson para proporcionar estimados más precisos de puntos percentiles $X_{0,00135}$, $X_{0,5}$ y $X_{0,99865}$ cuando los datos del proceso siguen una distribución no normal.

El concepto fundamental de este método se basa en la distribución normal. Generalmente, en los índices de capacidad de procesos se está interesado, dentro de la distribución del proceso, en tres puntos; es decir, la cola inferior, el punto de tendencia central y la cola superior. En términos de cantidades, estos puntos para la distribución normal corresponden, respectivamente a.

1. Cola superior: $X_{0,99865} = \mu + 3\sigma$
2. Cola inferior: $X_{0,00135} = \mu - 3\sigma$
3. Media= tendencia central: $X_{0,5} = \mu$

Cuando los datos se distribuyen normalmente, es sencillo estimar estos tres puntos. Sin embargo, en el caso cuando los datos no se distribuyen normalmente, no es fácil estimar estas cantidades, sobre todo cuando no se conoce la distribución de la variable fundamental. Más importantes, estas cantidades no corresponden a $\mu + 3\sigma$, μ y $\mu - 3\sigma$. Por ejemplo, el cuantil $X_{0,5}$ corresponde a la media μ para el caso normal, pero en el caso no normal, corresponde a la mediana.

Clements, utilizó el mismo enfoque para estimar cuantiles no normales y reemplazó 6σ en la ecuación 1.1 por las longitudes del intervalo entre los percentiles 0,135 y 99,865 de una determinada distribución. Por eso, el denominador de la ecuación (1.1) se puede reemplazar por $U_p - L_p$, es decir;

$$C_p = \frac{LES - LEI}{U_p - L_p} \quad (2.2)$$

donde, U_p es el percentil superior (99,865 percentil de observaciones) y L_p es el percentil (0,135 percentil de observaciones). La mediana “M” es el valor central preferido para una distribución asimétrica, así que los estimados de Clements C_{pi} y C_{ps} son como sigue.

$$C_{pi} = \frac{M - LEI}{M - L_p} \quad (2.3)$$

$$C_{ps} = \frac{LES - M}{U_p - M} \quad (2.4)$$

$$C_{pk} = \min(C_{ps}, C_{pi}) \quad (2.5)$$

El enfoque de Clements, no requiere una transformación matemática de los datos. Los profesionales no estadísticos pueden entenderlo y no requiere el ajuste de distribuciones complicadas (Kotz y Lovelace, 1998). Sin embargo, el método de Clements requiere el conocimiento de los términos de asimetría y kurtosis, los cuales se basan en el tercero y cuarto momento respectivamente, y estos pueden ser estimados no confiables para tamaños de muestras muy pequeños (Liu y Cheng, 2006). Wu et al. (1998), han conducido varios

estudios de investigación indicando que el método de Clements no mide exactamente los índices de capacidad, especialmente cuando la distribución de los datos es asimétrica.

2.3.2. Método de percentiles de Burr

Burr (1942), propuso una distribución llamada “Distribución Burr XII” para obtener los percentiles requeridos de una variable X. Este método es aplicable actualmente donde la robótica controla gran cantidad del proceso.

La función de densidad de probabilidad para una variable X, distribuida Burr XII se expresa como sigue.

$$f(x|c,k) = \begin{cases} ckx^{c-1}(1+x^c)^{-(k+1)} & \text{si } x \geq 0, c \geq 1, k \geq 1 \\ 0 & \text{si } x < 0 \end{cases} \quad (2.6)$$

donde, c y k representan los coeficientes de asimetría y la kurtosis de la distribución Burr. Por otro lado, la función de distribución acumulada y la función de transformación inversa de la distribución Burr XII están dadas respectivamente por.

$$F(x|c,k) = \begin{cases} 1 - (1+x^c)^{-k} & \text{si } x \geq 0, c \geq 1, k \geq 1 \\ 0 & \text{si } x < 0 \end{cases} \quad (2.7)$$

$$G(x|c,k) = \begin{cases} (1+x^c)^{-k} & \text{si } x \geq 0, c \geq 1, k \geq 1 \\ 0 & \text{si } x < 0 \end{cases} \quad (2.8)$$

Burr (1942) y (1973), logró tabular los valores esperados de la desviación estándar, el coeficiente de asimetría y kurtosis de esta distribución para diferentes combinaciones de c y k, es decir, se emplean los primeros cuatro momentos de una distribución, donde el tercer y cuarto momento de la distribución empírica considerada, se denotan (s_3 y s_4) y están dados por las siguientes expresiones matemáticas.

$$s_3 = \frac{n}{(n-1)(n-2)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad (2.9)$$

$$s_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n+1)^2}{(n-2)(n-3)} \quad (2.10)$$

donde, \bar{x} es la media muestral y s , es la desviación estándar muestral.

Los valores tabulados de asimetría y kurtosis representan una transformación estandarizada

entre una variable Burr, U , y otra variable aleatoria X , la cual está dada por, $\frac{X - \bar{x}}{s_x} = \frac{U - \mu}{\sigma}$

donde, \bar{x} , s_x , representan la media y la desviación estándar de la muestra original. Así mismo, U es la variable Burr seleccionada, μ y σ representan su media y desviación estándar.

Por ejemplo, una serie de datos con asimetría, $s_3=1,5$ y kurtosis, $s_4=6,2$ sigue una distribución Burr con parámetros $c=6,322081$ y $k=0,059357$ y con media y desviación respectivamente de $0,288250$ y $0,329972$.

Finalmente, Liu y Chen (2006), introdujeron una modificación al método de Clements, que utiliza percentiles en una familia de curvas de Pearson. En este método se sustituyen estos puntos percentiles por percentiles apropiados obtenidos por la distribución de Burr. Este método se resume en los siguientes pasos para una variable Burr U y una variable aleatoria X .

Paso 1: Se estiman la media (\bar{x}), desviación estándar, los puntos percentiles de la distribución empírica $X_{0,00135}$, $X_{0,50}$, $X_{0,99865}$, los coeficientes de asimetría y kurtosis (s_3) y (s_4) del conjunto de datos con distribución X dada.

Paso 2: Se seleccionan los parámetros c y k basados en los valores estimados de asimetría y kurtosis obtenidos en el paso anterior usando la distribución Burr XII, los cuales están

tabulados en Liu y Chen (2006) y cuyas expresiones son las siguientes; $c = \frac{(n-2)}{\sqrt{n(n-1)}} s_3$ y

$$k = \frac{(n-2)(n-3)}{(n^2-1)} s_4 + 3 \frac{(n-1)}{(n+1)}$$

Paso 3: Con esta misma tabla y de acuerdo a estos parámetros c y k obtenidos en el paso anterior, se obtienen los valores de los percentiles estandarizados $0,00135$; $0,5$ y $0,99865$ de la distribución Burr XII, estos percentiles se denotan como $Z_{0,00135}$, $Z_{0,5}$ y $Z_{0,99865}$ respectivamente.

Paso 4: Se estiman los percentiles inferior, mediana y superior, denotados L_p , M y U_p de la siguiente manera.

$$L_p = X_{0,00135} = \bar{x} + s_x Z_{0,00135} \quad (2.11)$$

$$M = X_{0,50} = \bar{x} + s_x Z_{0,50} \quad (2.12)$$

$$U_p = X_{0,99865} = \bar{x} + s_x Z_{0,99865} \quad (2.13)$$

Paso 5: Finalmente se estiman los índices de capacidad de procesos mediante las expresiones matemáticas dadas por (2.2) hasta (2.5).

2.3.3. Método Exacto

Este método permite también evaluar el desempeño de un proceso mediante los ICP. Czarski (2008), presenta una relación entre los percentiles $X_{0,00135}$, $X_{0,5}$ y $X_{0,99865}$, los cuales se definen a continuación.

$$\int_{-\infty}^{x_{0,00135}} f(x) dx = 0,00135 \quad (2.14)$$

$$\int_{-\infty}^{x_{0,50}} f(x) dx = 0,50 \quad (2.15)$$

$$\int_{-\infty}^{x_{0,99865}} f(x) dx = 0,99865 \quad (2.16)$$

Para determinar los valores de los percentiles usando las fórmulas (2.14)-(2.16) se requiere el conocimiento de la función de densidad de la distribución empírica. De manera general,

$$\int_{-\infty}^{x_i} f(x) dx = i, \text{ con } i = \{0,00135; 0,50; 0,99865\}. \text{ Pero } \int_{-\infty}^{x_i} f(x) dx := F(x)$$

2.3.4. Método ICP basado en la función de distribución de probabilidad.

Para $p \in [0,1]$, el p -ésimo cuantil de la función cuya fda es F_X , se define como el menor valor x_p que satisface la desigualdad, $p \leq F_X(x_p)$.

Obsérvese que por la definición de la función de distribución acumulada inversa se puede expresar que $x_p = F_X^{-1}(p) = \min\{x: p \leq F_X(x)\}$.

Cuando F_X es estrictamente creciente y continua, $F_X^{-1}(p)$ es el único valor de x_p que satisface,

$$F_X(x_p) = p \quad (2.17)$$

Cuando F_X no es estrictamente creciente o continua (como en el caso de que X sea discreta), puede existir más de una solución o posiblemente ninguna.

- Estimación de los percentiles.

Una manera de realizar una estimación natural de un cuantil poblacional $x_p = F_X^{-1}(p)$, es hacer $\hat{x}_p = \hat{F}_X^{-1}(p)$. Sin embargo, se debe tener presente que \hat{F}_X no es continuo y por tanto puede o no existir una solución a (2.17).

En primer lugar se ordenan los valores muestrales observados (x_1, x_2, \dots, x_n) para obtener el parámetro estadístico de orden $x_{(1)} < x_{(2)} \dots < x_{(n)}$. Observe ahora que $x_{(i)}$ es el (i/n) -ésimo cuantil de la distribución empírica, porque, $\hat{F}_X(x_{(i)}) = \frac{i}{n}$ o $\hat{F}_X(x_{(i)}) < \frac{i}{n}$; siempre que, $x < x_{(i)}$.

Generalizando el p -ésimo cuantil muestral es $\hat{x}_p = x_{(i)}$, cuando

$$\frac{i-1}{n} < p \leq \frac{i}{n} \quad (2.18)$$

Si se encuentra el valor de i , tal que (2.18) se cumple, entonces se hace;

$$\tilde{x}_p = x_{i-1} + n(x_{(i)} - x_{(i-1)}) \left(p - \frac{i-1}{n} \right) \quad (2.19)$$

donde, \tilde{x}_p es una interpolación lineal entre $x_{(i-1)}$ y $x_{(i)}$. Cuando n es par esta expresión da la mediana muestral $\tilde{x}_p = x_{(n/2)}$.

2.3.5. Transformación de potencia de Box-Cox

En este método, Box y Cox (1964), proponen una familia de transformaciones sobre una variable respuesta necesariamente positiva dada por.

$$X^{(\lambda)} = \begin{cases} \frac{X^\lambda - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \text{Ln}X & \text{para } \lambda = 0 \end{cases} \quad (2.20)$$

Esta familia de transformaciones continuas depende de un solo parámetro λ , el cual puede estimarse usando el método de máxima verosimilitud.

Para utilizar este método, Primero se selecciona un rango para el valor de λ , el cual se evalúa de la siguiente manera, resultando las ecuaciones que siguen.

$$\begin{aligned} L_{\max} &= -\frac{1}{2} \ln \hat{\sigma}^2 + \ln J(\lambda, X) = -\frac{1}{2} \ln \hat{\sigma}^2 + \ln \left[\prod_{i=1}^n \frac{\partial w_i}{\partial X_i} \right] \\ &= -\frac{1}{2} \ln \hat{\sigma}^2 + \ln \left[\prod_{i=1}^n X_i^{\lambda-1} \right] = -\frac{1}{2} \ln \hat{\sigma}^2 + (\lambda - 1) \ln \left[\prod_{i=1}^n X_i \right] \\ &= -\frac{1}{2} \ln \hat{\sigma}^2 + (\lambda - 1) \sum_{i=1}^n \ln X_i \end{aligned}$$

El valor $\frac{\partial w_i}{\partial X_i}$ representa la derivada parcial de cada función y $J(\lambda, X)$ el Jacobiano.

El estimador de $\hat{\sigma}^2$ para un valor fijo λ , está dado por, $\hat{\sigma}^2 = \frac{S(\lambda)}{n}$, donde $S(\lambda)$ es la suma de cuadrados residual en el análisis de varianza de $X^{(\lambda)}$.

Segundo, se calcula el valor $L_{\max}(\lambda)$ para distintos valores de λ dentro del rango establecido. El estimador de máxima verosimilitud de λ se obtiene por el valor que maximiza $L_{\max}(\lambda)$. Con el valor óptimo de λ^* cada ítem dentro de los límites de especificación es transformado a una variable normal. Los ICP son calculados finalmente

con la media y desviación estándar de los datos transformados usando las ecuaciones (2.2) - (2.5).

2.3.6. Transformación de Johnson

Este método de Johnson (1949), desarrolla un sistema de distribución basada en los métodos de los momentos, similar a un sistema de Pearson. La forma general de la transformación está dada por lo siguiente.

$$Z = \gamma + \eta\tau(x; \varepsilon, \lambda) \quad (2.21)$$

Los parámetros están sujetos a; $\lambda > 0$, $\eta > 0$, $-\infty < \varepsilon < \infty$, $-\infty < \gamma < \infty$ y además, z es la variable normal estándar. Por otro lado, x es la variable que debe ser estimada por una distribución de Johnson. Los cuatro parámetros γ , η , ε y λ pueden ser estimados, mientras que τ es una función arbitraria, la cual puede tomar alguna de las formas siguientes.

2.3.6.1. Sistema de Transformación Log-normal (S_L)

La distribución de Johnson formada por una familia log-normal, S_L , está dada por la siguiente expresión.

$$Z = \gamma + \eta \ln(x - \varepsilon) \quad (2.22)$$

con $x \geq \varepsilon$. Los estimadores para los parámetros están dados por.

$$\hat{\eta} = 1,645 \left[\log \left(\frac{x_{0,95} - x_{0,5}}{x_{0,5} - x_{0,05}} \right) \right]^{-1} \quad (2.23)$$

$$\hat{\gamma}^* = \hat{\eta} \log \left(\frac{1 - \exp\left(\frac{1,645}{\hat{\eta}}\right)}{x_{0,5} - x_{0,05}} \right) \quad (2.24)$$

$$\varepsilon = x_{0,5} - \exp\left(\frac{\hat{\gamma}^*}{\hat{\eta}}\right) \quad (2.25)$$

donde, el $\alpha(100)$ -ésimo percentil es obtenido como el valor $\alpha(n+1)$ -ésimo ranqueado de las n observaciones. Una observación importante es que es necesaria una interpolación lineal entre valores consecutivos que pueden ser útiles para determinar el percentil respectivo.

2.3.6.2. Sistema no acotado (S_U)

En este caso el valor τ_2 está dado por una familia de curvas no acotadas cuya expresión está dada de la siguiente manera.

$$Z = \gamma + \eta \operatorname{senh}^{-1} \left(\frac{x - \varepsilon}{\lambda} \right) \quad (2.26)$$

donde $-\infty < x < \infty$, $\lambda > 0$, $\eta > 0$, $-\infty < \varepsilon < \infty$, $-\infty < \gamma < \infty$

Esta familia de curvas involucran entre otras las distribuciones t y normal. Para la estimación de estas distribuciones Hahn y Shapiro (1967), generaron tablas para la determinación de $\hat{\gamma}$ y $\hat{\eta}$ basados en los valores dados de asimetría y kurtosis.

2.3.6.3. Sistema acotado (S_R)

La familia de sistemas acotados S_R involucra una distribución acotada, entre ellas las distribuciones Gamma y Beta. La distribución puede ser acotada bien sea en su parte final superior (ε) o final inferior ($\varepsilon + \alpha$) o en ambas, esto conduce a las siguientes situaciones.

- Caso 1: Rango de variación conocido. Para este caso, donde los valores finales en ambos lados de la distribución son conocidos, los parámetros se obtienen de la siguiente manera.

$$\hat{\eta} = \frac{Z_{1-\alpha'} - Z_{\alpha}}{\log \left(\frac{(x_{1-\alpha'} - \varepsilon)(\varepsilon + \lambda - x_{\alpha})}{(x_{\alpha'} - \varepsilon)(\varepsilon + \lambda - x_{1-\alpha'})} \right)} \quad (2.27)$$

$$\hat{\gamma} = z_{1-\alpha'} - \hat{\eta} \log \left(\frac{(x_{1-\alpha'} - \varepsilon)}{(\varepsilon + \lambda - x_{1-\alpha'})} \right) \quad (2.28)$$

- Caso 2: Un valor final es conocido. En este caso una ecuación más se obtiene apareando la mediana de los datos, esto para complementar las ecuaciones (2.24) y (2.25), dicha ecuación está dada por lo siguiente.

$$\hat{\lambda} = (x_{0.5} - \varepsilon) \left\{ (x_{0.5} - \varepsilon)(x_{\alpha} - \varepsilon) + (x_{0.5} - \varepsilon)(x_{1-\alpha} - \varepsilon) - 2(x_{\alpha} - \varepsilon)(x_{1-\alpha} - \varepsilon) \right\} \\ \left\{ (x_{0.5} - \varepsilon)^2 - (x_{\alpha} - \varepsilon)(x_{1-\alpha} - \varepsilon) \right\}^{-1} \quad (2.29)$$

- Caso 3: Ningún punto final es conocido. En este caso, ningún punto final es conocido y solo cuatro valores percentiles son escogidos como los percentiles de la distribución normal estándar. Para este caso cuando $i=1,2,3$ y 4 , se tienen las siguientes ecuaciones no lineales.

$$Z_i = \hat{\gamma} + \hat{\eta} \ln \left(\frac{x_i - \varepsilon}{\varepsilon + \hat{\lambda} - x_i} \right) \quad (2.30)$$

Por otro lado, Hill et al. (1976), desarrollaron un algoritmo que permite obtener el primer cuarto momento de X para las distribuciones de familias dadas en los 3 casos anteriores, al mismo tiempo los ICP pueden estimarse usando las ecuaciones (2.2)-(2.5).

2.3.7. Varianza Ponderada

Choi y Bai (1997), propusieron este método de varianza ponderada para ajustar los ICP considerando los grados de asimetría de la población desconocida. Sea X una variable del proceso en cuestión, y sea P_X la probabilidad de que dicha variable sea menor o igual que la media del proceso, μ .

$$\text{Esto es,} \quad P_X = \frac{1}{n} \sum_{i=1}^n I(\bar{X} - X_i) \quad (2.31)$$

$$\text{donde, } I(X) = \begin{cases} 1 & \text{para } x > 0 \\ 0 & \text{para } x < 0 \end{cases}$$

Así los índices C_p y C_{pk} basados en este método se estiman mediante lo siguiente.

$$C_p = \frac{LES - LEI}{6\sigma W_X} \quad (2.32)$$

$$\text{donde, } W_X = \sqrt{1 + |1 - 2P_X|}.$$

Por otro lado se tiene.

$$C_{pu} = \frac{LES - \mu}{3\sqrt{2P_x}\sigma} \quad (2.33)$$

$$C_{pl} = \frac{\mu - LEI}{3\sqrt{2(1-P_x)}\sigma} \quad (2.34)$$

Entonces el índice C_{pk} está dado por, $C_{pk} = \min[C_{pl}, C_{pu}]$.

2.3.8 Índice de Capacidad de Procesos de Wright, C_s

Wright (1995), propuso un ICP considerando la asimetría e incorporando un factor de corrección de asimetría en el denominador del índice C_{pmk} , definiéndolo de la siguiente

forma, $C_s = \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{\sigma^2 + (\mu - T)^2 + \left|\frac{\mu_3}{\sigma}\right|}} = \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{\sigma^2 + \left|\frac{\mu_3}{\sigma}\right|}}$

donde, $T = \mu$ y μ_3 es el tercer momento central.

Algunos de los métodos antes descritos, se han utilizado en la industria. Sin embargo, el método de Transformación de Box Cox se utiliza muy poco.

En general, los índices de capacidad de procesos basados en estos métodos fueron propuestos para superar la deficiencia de no normalidad en algunos procesos. Por esta razón se trata de determinar su desempeño y confiabilidad.

Otra manera de estudiar ese desempeño es usando el método de bootstrap para estimar los intervalos de confianza para los ICP basados en las técnicas de Clements y de Burr, para posteriormente comparar entre sí los diferentes intervalos de confianza bootstrap como son; el estándar, el percentil y el percentil de sesgo corregido, los cuales se escriben a continuación.

2.3.9. Intervalos de Confianza para los estimados de ICP: C_p y C_{pk}

Bajo el supuesto de distribución normal es posible derivar expresiones útiles para las propiedades estadísticas de los estimados de ICP; desafortunadamente cuando el supuesto

de normalidad no se cumple, la derivación de resultados relevantes para el trabajo práctico se hace extremadamente difícil. A fin de evitar esta dificultad, la aplicación de los métodos de computación estadística han sido considerado; en particular Franklin y Wasserman (1992); Kotz y Johnson (1993), desarrollaron la técnica bootstrap, aunque fue propuesto años atrás como solución al problema por Efron y Tibshirani (1986). Esto representa el único enfoque viable en la determinación de intervalos de confianza para el C_{pk} no normal.

En el desarrollo de esta trabajo se construyen y comparan tres diferentes tipos de intervalos de confianzas, el estándar bootstrap, el percentil bootstrap y el percentil insesgado corregido bootstrap, denotados respectivamente SB, PB PSCB.

En la construcción de los intervalos presentados por Franklin y Wasserman (1992), la idea es construir una muestra bootstrap, $x_1^*, x_2^*, \dots, x_n^*$ de tamaño n con reemplazamiento, de una muestra original proveniente de un proceso, x_1, x_2, \dots, x_n . Se determinan valores de $\hat{C}_p^*, \hat{C}_{pk}^*, \hat{C}_{pm}^*$, los cuales son estimadores de la distribución de $\hat{C}_p, \hat{C}_{pk}, \hat{C}_{pm}$ respectivamente. Se estiman tres intervalos de confianza usando la técnica bootstrap y se consideran B estimadores bootstrap de cada índice C_p y C_{pk} ordenados desde de manera creciente. La notación usada por Franklin y Wasserman en su estudio es la siguiente, \hat{C} y $\hat{C}^*(i)$, para indicar el estimador del índice de capacidad y el estimador bootstrap ordenado asociado a cada índice. Por ejemplo, el valor más pequeño de las B muestras bootstrap de C_p y C_{pk} respectivamente los denota: $\hat{C}_p^*(1)$ y $\hat{C}_{pk}^*(1)$.

2.3.9.1 Estándar Bootstrap, (SB).

Para los B estimadores bootstrap $\hat{C}^*(i)$ se determina un promedio muestral y una desviación estándar muestral con las siguientes expresiones.

$$\bar{C}^* = \frac{1}{B} \sum_{i=1}^B \hat{C}^*(i) \quad (2.35)$$

$$S_c^* = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\hat{C}^*(i) - \bar{C}^*)^2} \quad (2.36)$$

El valor S_c^* es el estimador de la desviación estándar de \hat{C} , luego si la distribución de \hat{C} es aproximadamente normal, entonces el intervalo de confianza SB está dado de la siguiente manera:

$$\bar{C}^* \pm z_{\alpha} S_C^* \quad (2.37)$$

donde z_{α} es el percentil α de la distribución normal estándar.

2.3.9.2 Percentil Bootstrap, (PB).

Como se tomará una muestra bootstrap ordenada de los estimados de $\hat{C}^*(i)$, el intervalo PB, de $(1-2\alpha)100\%$ para C^* , se obtiene de la siguiente manera,

$$\left[\hat{C}^*(\alpha B), \hat{C}^*((1-\alpha)B) \right] \quad (2.38)$$

2.3.9.3 Percentil Sesgo-Corregido Bootstrap (PSCB)

Puede ocurrir que la distribución de la muestra bootstrap obtenida sea insesgada, por lo que se considera un estimado del insesgamiento, (Efron (1982)) el cual es la diferencia entre el valor de la media del bootstrap y el valor estimado inicial. Para construir el intervalo PSCB, primero se calcula la probabilidad, p_0 , usando la distribución ordenada de \hat{C}^* , esta probabilidad está dada por, $p_0 = \Pr[\hat{C}^* \neq \hat{c}]$, luego usando la distribución acumulativa normal estándar $\Phi(\cdot)$ se determinan los siguientes valores.

$$z_0 = \Phi^{-1}(p_0) \quad (2.39)$$

$$p_L = \Phi(2z_0 - z_{\alpha}) \quad (2.40)$$

$$p_U = \Phi(2z_0 + z_{\alpha}) \quad (2.41)$$

Finalmente, el intervalo PSCB, queda determinado de la siguiente manera.

$$\left[\hat{C}^*(P_L \cdot B), \hat{C}^*(P_U \cdot B) \right] \quad (2.42)$$

Estos intervalos fueron utilizados para evaluar los ICP a través de la distribución Pareto Generalizada con 2 parámetros.

CAPÍTULO 3

AJUSTE DE DISTRIBUCIONES CON DATOS PROVENIENTES DE VARIABLES DE CALIDAD NO NORMALES

3.1. INTRODUCCION

En este capítulo se describe un enfoque simple para resolver el problema de estimación de ICP no normales. Este tema trata con ajuste de distribuciones conocidas a datos provenientes de una variable de calidad distribuida no normalmente. También se hace referencia a un algoritmo genético utilizado para estimar los parámetros de la distribución Pareto Generalizada con 3 parámetros (DPG3), el cual se emplea en el capítulo 5.

3.2 AJUSTE DE DISTRIBUCIONES CONOCIDAS CON VARIABLES DE CALIDAD NO NORMALES

Aunque la estimación de percentiles y las técnicas de transformación de datos se utilizan comúnmente para la resolución de los problemas de no normalidad en algunas variables de calidad, existe otro enfoque para manejar este tópico. De acuerdo a esto, algunas distribuciones como Gamma, Weibull, Beta y Log-normal se ajusta a los datos muestrales provenientes de variables de calidad (Sommerville y Montgomery, 1997). Entonces, los índices de capacidad de procesos se pueden evaluar usando el porcentaje que caen fuera de los límites de especificación de la distribución ajustada. Hahn y Shapiro (1997), sugirieron usar la asimetría y kurtosis muestral calculada y graficándola, la cual puede dar una idea de la distribución apropiada para uso del ajuste.

Ajustar las distribuciones conocidas con datos de variables de calidad, se puede llevar a cabo usando diversos software estadísticos. El analista puede determinar cuál distribución puede ajustar “mejor” a los datos muestrales. Consecuentemente, puede desplegar técnicas directamente para evaluar los índices de capacidad de procesos. Ajustar una distribución conocida a datos muestrales tiene diversas ventajas, es decir, un enfoque directo y el analista trata con la calidad de los datos muestrales en lugar de hacer ajustes y aproximaciones, al mismo tiempo que puede agudizar su ingenio en el proceso cuando se ajustan diferentes distribuciones. Este ingenio podría guiarlo en la implementación de

intentos de mejoramiento al proceso real. El defecto de este enfoque es que una muestra relativamente grande se debe obtener de tal manera que el ajuste de la distribución se pueda realizar con algún grado de confianza (Somerville y Montgomery, 1997).

Es de hacer notar que distintos enfoques se han propuesto para resolver el problema que ocasiona trabajar con datos distribuidos no normales. Muchos trabajos se han encaminado en la modificación de los ICP tradicionales tomando en cuenta que los datos en estudio siguen una distribución no normal. Spiring et al. (2003), hace una recopilación de artículos con este tipo de caso, allí se recogen diferentes bibliografías de autores cuyos enfoques son los ICP no normales.

A continuación se presentan las funciones de densidad de las distribuciones Weibull, Log-normal y Beta. Además, se presenta una descripción detallada de la DPG con 2 y 3 parámetros, ya que son los modelos a utilizar en este trabajo. Además se presenta la DPG con 4 parámetros, la cual fue presentada teóricamente por Elfattah, et al. (2007). En las conclusiones finales, se sugiere el uso de este modelo para futuras aplicaciones.

Las distribuciones Weibull, Beta, Log-normal y Gamma, se usan para investigar el efecto de datos no normales en la estimación de índices de capacidad de procesos. Se sabe que estas distribuciones tienen valores paramétricos que pueden representar desviaciones moderadas a severas de la normalidad. Los parámetros se seleccionan de tal manera, que puede compararse con otros resultados expuestos por otros autores como por ejemplo Tang y Than (1999) o Liu y Chen (2006). A continuación se presentan las funciones de densidad de probabilidad de las distribuciones Weibull, Beta, Gamma y Log-normal. Seguidamente, se describe la Distribución Pareto Generalizada, la cual fue considerada en el desarrollo del trabajo como modelo de ajuste en la estimación de los índices de capacidad de procesos.

- **Distribución Weibull**

Se dice que una variable aleatoria X , tiene una distribución de Weibull, si su función de densidad de probabilidad con parámetros de forma α y de escala β , está dada por:

$$f(x, \alpha, \beta) = \begin{cases} \frac{\alpha}{\beta^\alpha} (x^{\alpha-1}) \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right] & \text{para } x \neq 0; \alpha, \beta > 0 \\ 0 & \text{para cualquier otro valor} \end{cases} \quad (3.1)$$

El η -ésimo cuantil de esta distribución está dado de la siguiente manera, $x(\alpha, \beta) = \beta [-\ln(1-\eta)]^{\alpha-1}$ con $x \geq 0$.

Por otro lado, la función Log-verosimilitud para una muestra distribuida Weibull x_1, x_2, \dots, x_n está representada como; $l(\alpha, \beta) = n \ln\left(\frac{\alpha}{\beta}\right) + (\alpha-1) \sum_{i=1}^n \ln x_i - n(\alpha-1) \ln \beta - \sum_{i=1}^n \left(\frac{x_i}{\beta}\right)^\alpha$

- **Distribución Log-normal**

Sea X una variable aleatoria distribuida normalmente con media μ y varianza σ^2 . La función de densidad de probabilidad para una variable de la forma, $Y = \exp(X)$, está dada de la siguiente manera:

$$f_Y(y, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma y}} \exp\left\{-\frac{1}{2} \left[\frac{\ln(y) - \mu}{\sigma}\right]^2\right\} \quad \text{para } y > 0 \quad (3.2)$$

con, $-\infty < \mu < \infty$ y $x \geq 0$. El η -ésimo cuantil de esta distribución se estima como, $x(\alpha, \beta) = \exp\left[\mu + \Phi^{-1}(\eta)\sigma\right]$, donde $\Phi(\cdot)$ es la función de densidad acumulativa de la distribución normal.

Para una muestra aleatoria distribuida log-normal x_1, x_2, \dots, x_n se tiene la función log-verosimilitud de μ y σ ; $\ln L(\theta) = \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \sum_{i=1}^n \ln x_i - \frac{1}{2} \sum_{i=1}^n \left(\frac{\ln x_i - \mu}{\sigma}\right)^2$.

- **Distribución Beta**

Sea X una variable aleatoria, se dice que X tiene una distribución Beta, si la función de densidad de probabilidad está dada por:

$$f(x, \alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{para } 0 < x < 1; \alpha, \beta > 0 \\ 0 & \text{para cualquier otro valor} \end{cases} \quad (3.3)$$

- **Distribución Pareto Generalizada (DPG)**

La distribución de Pareto fue propuesta en 1897 por Wilfred Pareto, como un modelo para distribuciones de ingreso y hoy en día se sigue utilizando para modelar diferentes situaciones, específicamente en datos de lluvia. Puede ser utilizada además para representar otras distribuciones, de hecho es una combinación de distribución exponencial y gamma. Juega un papel muy importante en el estudio sobre el tamaño de poblaciones en ciudades, acontecimientos que involucren riquezas naturales, riesgo en seguros y fallas comerciales. Arnald et al. (1983), presentaron un resumen sobre el historial de uso de dicha distribución en el contexto de distribución de ingresos.

Posteriormente, la Distribución Pareto Generalizada, DPG, fue presentada por Pickands (1975), y desde entonces su aplicabilidad ha sido utilizada en muchas áreas sobre todo en el campo de la economía.

Embrechts et al. (1997), la introdujo en el contenido de la teoría de funciones de valores extremos. La función de distribución de Pareto Generalizada, está dada de la siguiente manera:

$$F(x, \alpha) = \begin{cases} 1 - (1 + \alpha x)^{-\frac{1}{\alpha}} & \text{si } \alpha \neq 0 \\ 1 - \exp(-x) & \text{si } \alpha = 0 \end{cases} \quad (3.4)$$

donde, $x \geq 0$ si $\alpha \geq 0$ y $x \in [0, -1/\alpha]$ en otro caso.

La función de densidad está dada como: $f(x, \alpha) = \frac{\alpha}{x^{\alpha+1}}$, para $x=1$ y $c < 0$, donde α es el parámetro de forma.

Cuando se introduce en el modelo anterior otro parámetro, se tiene el modelo para la distribución Pareto Generalizada con 2 parámetros (DPG2). Entre las

aplicaciones de la DPG2 se considera análisis de funciones de valor extremo, modelación de fenómenos socioeconómicos, procesos hidrológicos, procesos físicos y biológicos, análisis de medio ambiente, distribuciones para tiempos de fallas, situaciones en las cuales la distribución exponencial se puede utilizar pero se requiere algún estudio de robustez en alternativa de colas. Davinson y Smith (1990), puntualizaron que la DPG2 es la base de todo un enfoque extenso de modelación con un alto nivel de excelencia. Du-Mouchel (1983), la aplicó para estimar el índice α de estabilidad para medidas de colas. Davison (1984a y 1984b), modeló la contaminación atmosférica producida por el transporte de radionuclides de rango largo. Van Montfort y Witter (1985 y 1986), Van Montfort y Otten (1991), la aplicaron para modelar los puntos máximos sobre un valor umbral (peak over a threshold, (POT)) para una serie de flujo de corrientes y lluvias. Smith (1984, 1987 y 1991), la utilizó en el análisis de inundaciones frecuentes y ondas.

Rosbjerg et al. (1992), compararon el uso de una distribución PG con 2-parámetros y una distribución exponencial como modelos de distribución para excedencias considerando una distribución de origen proveniente de una distribución PG.

Una versatilidad que presenta la DPG2 es que bien puede transformarse de acuerdo a los valores asignados a los parámetros, en otras distribuciones conocidas. Por ejemplo una $DPG(-1,414; 0,322)$ es equivalente a una Weibull(0,5); una $DPG(-0,14;1,406)$ es equivalente a una Log-normal o una $DPG(0,251; 2,455)$ es equivalente a una distribución Gamma(2,1).

Por otro lado, Öztekin (2005), definió para una variable aleatoria Y con distribución exponencial estándar, una Distribución Pareto Generalizada con 3 parámetros. Se agrega un nuevo parámetro denominado “parámetro de posición” el cual se define más adelante, así mismo sus propiedades.

En la revisión de la literatura realizada hasta ahora no se ha logrado encontrar la aplicación de la DPG en el área de control estadístico de procesos, es por eso que en este trabajo se introduce este modelo con 2 parámetros aplicado en la industria.

- **Distribución Pareto Generalizada con 2 parámetros**

La distribución Pareto Generalizada se puede definir en función de la cantidad de parámetros involucrados en la función de distribución acumulativa, PDF.

Hosking y Wallis (1987), presentan una definición de la DPG, tomando en cuenta una variable aleatoria X , definida de la manera siguiente, $X = \frac{\beta(1 - e^{-\alpha Y})}{\alpha}$, donde Y es una variable aleatoria con distribución exponencial, donde β y α son los parámetros. La distribución de X se dice que es Pareto Generalizada con dos parámetros, si la función de distribución acumulativa está definida de la siguiente manera:

$$F(x, \alpha, \beta) = \begin{cases} 1 - \left(1 - \frac{\alpha x}{\beta}\right)^{\frac{1}{\alpha}} & \text{si } \alpha \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{si } \alpha = 0 \end{cases} \quad (3.5)$$

Por otro lado, la función de densidad de probabilidad (fdp) y la función de estimación de los percentiles de la DPG2 están dadas de la siguiente manera;

$$f(x, \alpha, \beta) = \begin{cases} \beta^{-1} \left(1 - \frac{\alpha x}{\beta}\right)^{\frac{1-\alpha}{\alpha}} & \text{si } \alpha \neq 0 \\ \beta^{-1} \exp\left(-\frac{x}{\beta}\right) & \text{si } \alpha = 0 \end{cases} \quad (3.6)$$

$$F_p^{-1} = x(F, \alpha, \beta) = \begin{cases} \frac{\beta}{\alpha} \{1 - (1-F)^\alpha\} & \text{si } \alpha \neq 0 \\ -\beta \log(1-F) & \text{si } \alpha = 0 \end{cases} \quad (3.7)$$

El rango de x es $0 \leq x < \infty$ para valores de $\alpha \leq 0$; además para $\beta > 0$ se tiene un rango de $0 \leq x < \beta/\alpha$. Los parámetros de la DPG son: β , el parámetro de escala y α , el parámetro de forma. Un estimador $\hat{x}(F)$ se define sustituyendo los valores de $\hat{\beta}$ y $\hat{\alpha}$ para los valores obtenidos en la ecuación de $x(F, \alpha, \beta)$. Cuando $\alpha=0$, se tiene la

distribución exponencial con media β y cuando $\alpha=1$ se tiene la distribución uniforme sobre $[0,\beta]$. Cuando $\alpha<0$, esta distribución se denomina Pareto; finalmente para $\alpha>1/2$, la DPG presenta finitos puntos para $f(x)>0$.

Los valores de la media, varianza, asimetría y kurtosis pueden estimarse mediante las expresiones siguientes.

1. Media, $\frac{\beta}{1+\alpha}$

2. Varianza, $\frac{\beta^2}{[(1+\alpha)^2(1+2\alpha)]}$

3. Asimetría, $\frac{2(1-\alpha)(1+2\alpha)^{1/2}}{(1+3\alpha)}$

4. Kurtosis, $\frac{3(1+2\alpha)(3-\alpha+2\alpha^2)}{(1+3\alpha)(1+4\alpha)} - 3$

En las figuras 3.1 y 3.2 se presentan la función de densidad de probabilidad de una familia de distribuciones PG2 para diferentes valores del parámetro α y fijando $\beta=1$.

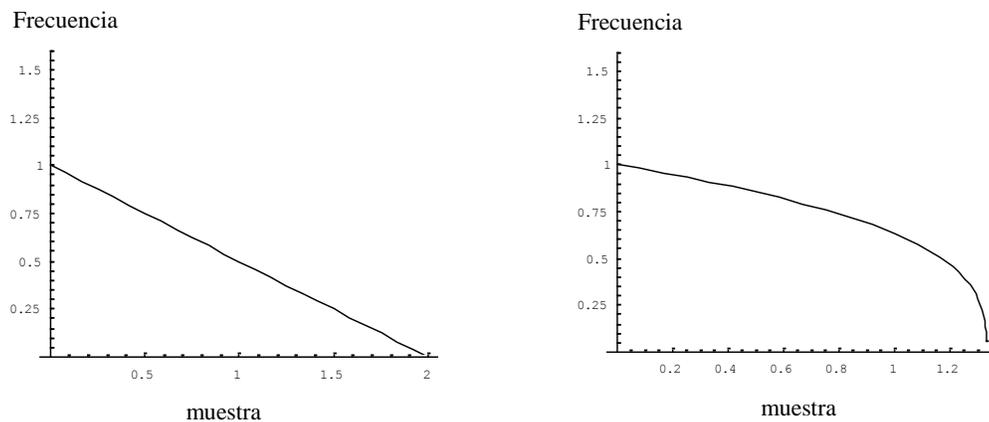


Figura 3.1: Función de densidad de probabilidad de la DPG2, para $\beta=0,5$ y $\alpha=0,75$

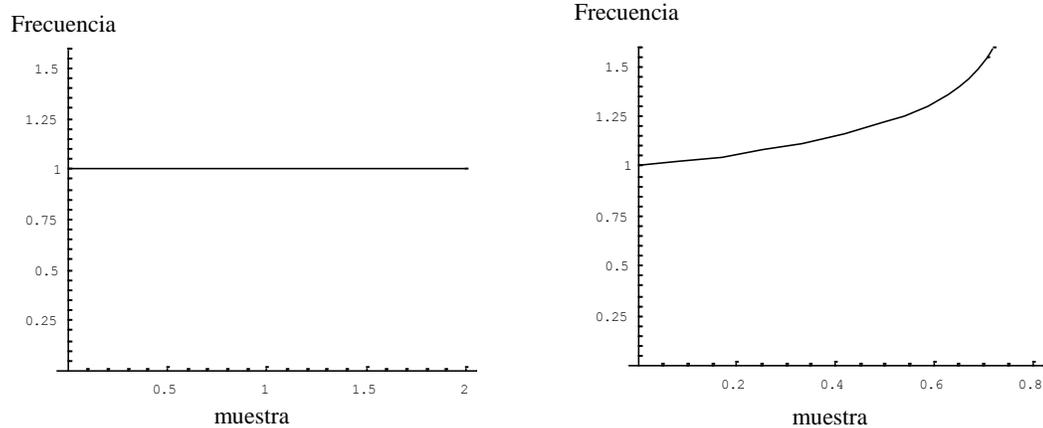


Figura 3.2: Función de densidad de probabilidad de la DPG, para $\beta=1,0$ y $\alpha=1,25$.

3.3 ALGORITMO GENETICO.

Un algoritmo es una serie de pasos organizados que describe el proceso que se debe seguir, para dar solución a un problema específico. John Henry Holland (1975) y DeJong (1975), promovieron una de las líneas más prometedoras de la inteligencia artificial como lo son los algoritmos genéticos (AG), llamados así porque se inspiran en la evolución biológica y su base genético-molecular. Estos algoritmos hacen evolucionar una población de individuos sometiéndola a acciones aleatorias semejantes a las que actúan en la evolución biológica (mutaciones y recombinaciones genéticas), así como también a una selección de acuerdo con algún criterio, en función del cual se decide cuáles son los individuos más adaptados, que sobreviven, y cuáles son menos aptos al mismo tiempo que son descartados. Es un método de búsqueda dirigida basado en probabilidad. Bajo una condición muy débil (que el algoritmo mantenga elitismo, es decir, guarde siempre al mejor elemento de la población sin hacerle ningún cambio), se puede demostrar que el algoritmo converge en probabilidad al óptimo. En otras palabras, al aumentar el número de iteraciones, la probabilidad de tener el óptimo en la población tiende a uno. (Gulsen, et al. (1995)).

Los algoritmos genéticos se incluyen dentro de los algoritmos evolutivos, que incluyen también las estrategias evolutivas, la programación evolutiva y la programación genética. Generalmente son utilizados en la optimización de problemas estocásticos considerando condiciones iniciales, basados en leyes biológicas de selección natural y genes. La diferencia con los métodos de búsquedas tradicionales es que, estos comienzan con una

solución inicial seleccionada aleatoriamente y la cual se le denomina población (Goldber, (1989)). En economía se utilizan para la estimación de los parámetros de modelos econométricos no lineales (Öztürkler y Altan (2008)). Gen y Chen (1996), demostraron que los algoritmos genéticos producen mejores resultados que los métodos tradicionales de optimización. En Rivas et al. (2006), se presenta una aplicación de algoritmos genéticos, donde se estiman los parámetros de un modelo de regresión de Cox en análisis de supervivencia, allí se muestra mediante una aplicación, que efectivamente estos algoritmos generan mejores resultados en la estimación de dichos parámetros. Así mismo, hacen referencia al texto presentado por Aguilar y Rivas (2001), donde es posible obtener información acerca de las técnicas de computación evolutiva necesarias para el entendimiento del tema.

Los algoritmos genéticos son útiles en la estimación de parámetros de modelos no lineales, cuyas funciones resultan áridas de trabajar; en este trabajo se utilizó esta herramienta con el objetivo de estimar los parámetros que conforman a la DPG3; esto es, se propuso un algoritmo genético que permitió estimar los parámetros del modelo de la DPG3 considerando valores iniciales obtenidos con un método tradicional. Una vez obtenidos estos valores estimados de los parámetros, mediante un algoritmo genético, se simuló muestras con las cuales se determinaron finalmente los ICP no normales mediante los métodos de percentiles de Clements, percentiles de Burr y método exacto.

No obstante, aunque los algoritmos genéticos producen resultados óptimos en comparación con otros métodos, el objetivo en esta sección del trabajo se basó en la estimación de los ICP no normales mediante el modelo de la DPG3, pero se hace hincapié en que los resultados obtenidos de los parámetros estimados mediante la técnica de algoritmos genéticos son más próximos a los valores iniciales de dichos parámetros en comparación con el programa elaborado usando el software R.

3.3.1. Funcionamiento

Como ya se dijo anteriormente, los algoritmos genéticos pueden generar un conjunto de soluciones de un problema, el cual es llamado fenotipo; mientras que el conjunto de

individuos de una población natural (codificando la información de cada solución en una cadena generalmente binaria), se denominan cromosomas. Los símbolos que forman la cadena son llamados genes. La representación de los cromosomas se hace con cadenas de dígitos binarios conocidos como genotipos. Por ejemplo, un cromosoma puede estar representado de la siguiente manera:

Cromosoma=[10110, 11001, 11011]

Los cromosomas evolucionan a través de iteraciones denominadas generaciones. En cada generación, los cromosomas son evaluados usando alguna función de aptitud, ese procedimiento se denomina evaluación. Las generaciones que siguen son llamadas nuevos cromosomas, los cuales pueden ser cruzados o mutados.

3.3.2. Cuándo usar estos algoritmos

Los algoritmos genéticos son de probada eficacia en caso de querer calcular funciones no derivables (o de derivación muy compleja), aunque su uso es posible con cualquier función. Según, Gulsen et al. (1995), se deben tener en cuenta las siguientes consideraciones:

- Si la función que se va a optimizar tiene muchos máximos o mínimos locales, se requerirán también muchas iteraciones del algoritmo para asegurar el máximo y/o mínimo global.
- Si la función a optimizar contiene varios puntos muy cercanos en valor al óptimo, solamente se puede asegurar la existencia de ellos (no necesariamente el óptimo).

Chatterjee y Laudano (1997), establecen los siguientes problemas en los cuales el uso de algoritmos genéticos es común, se utilizan como herramientas para estimar parámetros en problemas determinísticos, problemas de mínimos cuadrados no lineales, mínimos cuadrados con una variable independiente, problemas de regresión múltiple, reducir la dimensión de un modelo de regresión para escoger el conjunto de posibles mejores soluciones.

3.3.3. Operación de un algoritmo genético.

Un algoritmo genético puede presentar diversas variaciones, dependiendo de cómo se aplican los operadores genéticos (cruzamiento, mutación), de cómo se realiza la selección y de cómo se decide el reemplazo de los individuos para formar la nueva población. Yerniay (1999), establece los siguientes pasos para resolver un problema mediante el uso de algoritmos genéticos.

Paso 1: Se escoge un conjunto de códigos compatibles con la naturaleza del problema, esto se denomina, inicialización; es decir, se genera aleatoriamente la población inicial, que está constituida por un conjunto de cromosomas los cuales representan las posibles soluciones del problema. En caso de no hacerla aleatoriamente, es importante garantizar que dentro de la población inicial, se tenga la diversidad estructural de estas soluciones para tener una representación de la mayor parte de la población posible o al menos evitar la convergencia prematura.

Paso 2: Se combinan los cromosomas obtenidos de la población inicial creando nuevas generaciones. Se evalúa a cada uno de los cromosomas obtenidos mediante una función de aptitud para saber qué tan "buena" es la solución que se está codificando.

Paso 3: Obtenida una nueva población y ya evaluada se fija la condición de finalización del algoritmo genético. El algoritmo genético se deberá detener cuando se alcance la solución óptima, pero ésta generalmente se desconoce, por lo que se deben utilizar otros criterios de detención. Normalmente se usan dos criterios: correr el algoritmo genético un número máximo de iteraciones (generaciones) o detenerlo cuando no haya cambios en la población. Mientras no se cumpla la condición de término se siguen los siguientes pasos:

- Selección: Después de saber la aptitud de cada cromosoma, se procede a elegir los cromosomas que serán cruzados en la siguiente generación. Los cromosomas con mejor aptitud tienen mayor probabilidad de ser seleccionados.
- Cruzamiento: La recombinación es el principal operador genético, representa la reproducción sexual, opera sobre dos cromosomas a la vez para generar dos descendientes donde se combinan las características de ambos cromosomas padres.

- **Mutación:** Se modifica al azar, parte del cromosoma de los individuos, esto permite alcanzar zonas del espacio de búsqueda que no estaban cubiertas por los individuos de la población actual.
- **Reemplazo:** Una vez aplicados los operadores genéticos, se seleccionan los mejores individuos para conformar la población de la generación siguiente.

En este trabajo se construyó un algoritmo genético que permitió estimar los parámetros del modelo de la distribución Pareto Generalizada con 3 parámetros y los resultados se compararon con los obtenidos con el software R. Esto permite pensar que ciertamente, esta herramienta de investigación puede ser utilizada en cualquier área donde se quiera conocer los parámetros de un modelo determinado. La aplicación del algoritmo genético se encuentra en el capítulo 5 y el programa está desarrollado en la sección de anexos (anexo 6).

www.bdigital.ula.ve

c.c Reconocimiento

CAPÍTULO 4

ANÁLISIS DE CAPACIDAD DE PROCESOS UNIVARIANTES. APLICACIONES

4.1 INTRODUCCION

En este capítulo se utilizan los métodos descritos en el capítulo 2 para estimar los índices de capacidad de procesos C_p y C_{pk} considerando distribuciones no normales, tales como: log-normal, Weibull, Gamma y Pareto Generalizada, todas con 2 parámetros. Se simulan diferentes muestras para cada caso, con el objetivo de ilustrar mediante histogramas el tipo de desviación de la normalidad generada por cada distribución. Es decir, una muestra simulada con distribución Log-normal con parámetros (0,5; 0,2) produce datos con desviación leve de la normalidad; pero otra muestra generada mediante una distribución Weibull con parámetros (1; 1,2) generan datos con desviación severa de la normalidad. Surge las interrogantes, ¿Cuál de estas dos distribuciones es la adecuada para estimar los ICP? ¿Qué métodos se debe seguir, el de percentil de Clements, percentil de Burr, el método exacto o el método tradicional?

Por otro lado, se presenta el método de máxima verosimilitud y de momentos para estimar los parámetros de una DPG2, se consideran datos reales donde se demuestra el ajuste del modelo PG2 a un conjunto de datos, de igual manera se simulan datos con dicha distribución para estimar los ICP y determinar su desempeño. Además de esto, otro método para estudiar el desempeño de los ICP que se toman en cuenta en este capítulo, son los métodos de estimación por intervalos bootstrap, los cuales se utilizan cuando la distribución de los datos es no normal y permiten determinar la capacidad de cubrimiento que tienen. Para ello, se toma como modelo la distribución Pareto Generalizada con 2 parámetros.

4.2 ESTIMACION DE CAPACIDAD DE PROCESOS USANDO DISTRIBUCIONES NO NORMALES. APLICACIÓN.

En esta sección del trabajo, se realiza un estudio comparativo del desempeño de los ICP no normales utilizando los métodos de Percentil de Clements, Percentil de Burr y el método tradicional asumiendo conocidos los parámetros de las distribuciones.

Entre las distribuciones consideradas para el estudio comparativo se encuentran: Weibull, Beta, Log-normal y Pareto Generalizada con parámetros conocidos. Estas, se utilizan para investigar el efecto que tiene la no normalidad en los datos al estimar los índices de capacidad de procesos. Cada una de ellas representa un grado de desviación de la normalidad que van desde moderadas hasta severas y cada una produce un efecto en los resultados de los ICP estimados. Con el fin de ilustrar la condición de desviación en la normalidad, se presentan las figuras 4.1, 4.2, 4.3 y 4.4.

Se escogen las distribuciones log-normal (0,5; 0,2) la cual genera una desviación leve en los datos simulados; Gamma (4,0 ; 0,5) que genera desviación moderada y Weibull(1,0; 1,2) que genera una desviación severa de la normalidad, esto para generar muestras de tamaño 1000 y verificar dicha condición.

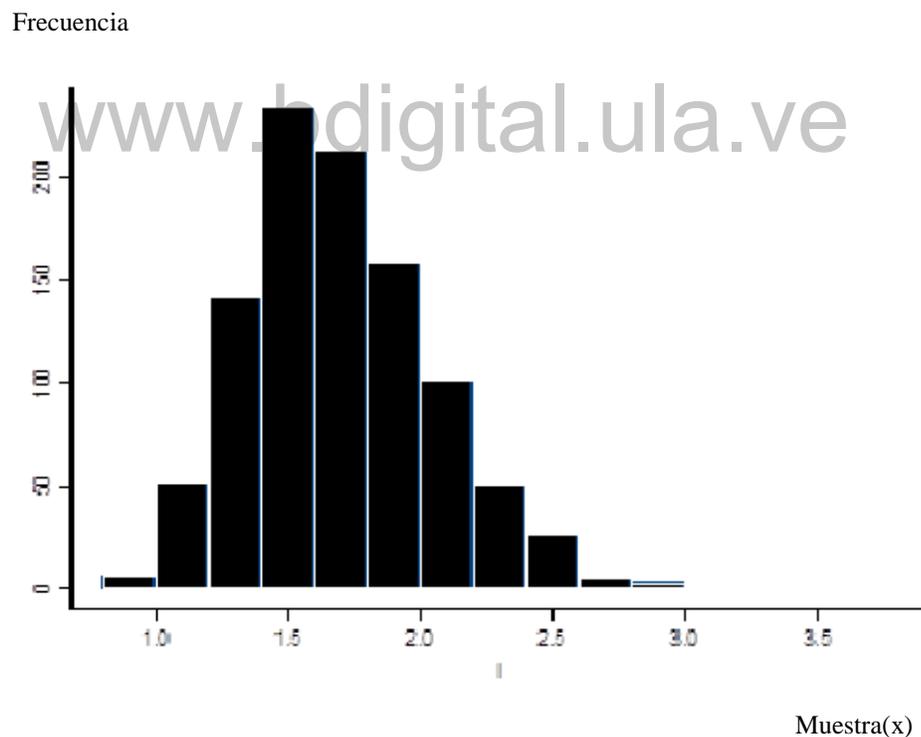


Figura 4.1. Distribución Log-normal(0,5; 0,2).

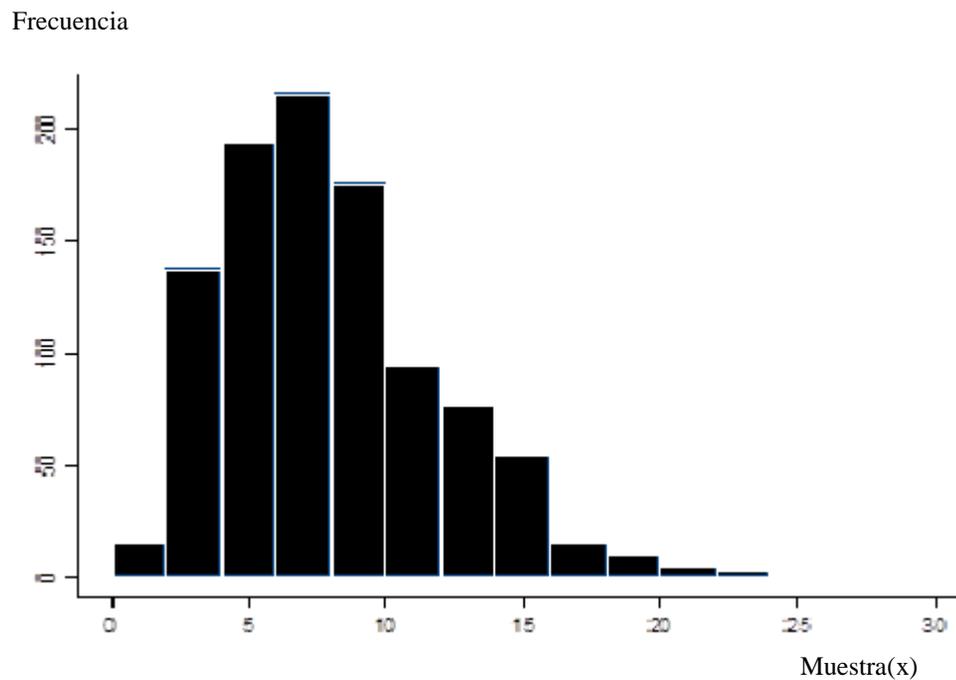


Figura 4.2. Distribución Gamma(4,0; 0,5)

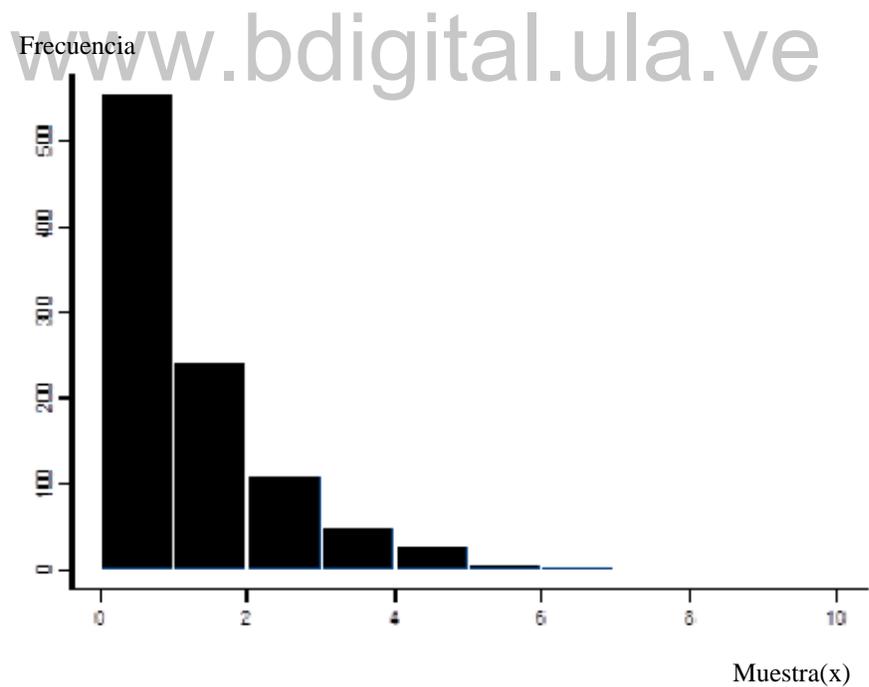


Figura 4.3. Distribución Weibull(1,0; 1,2)

De manera similar en la figura 4.4, se muestra una distribución Pareto Generalizada con 2 parámetros que representa desviación severa de normalidad.

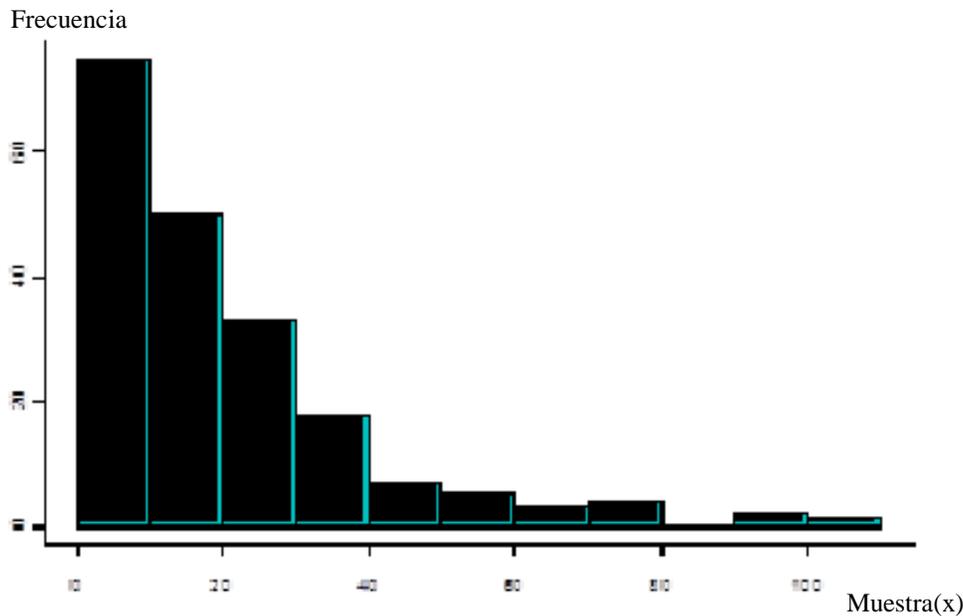


Figura 4.4. Distribución PG(20,7; 0,04)

4.2.1 Estudio de simulación

El objetivo principal es estimar los ICP para evaluar su desempeño cuando las distribuciones de estudio no son normales. Para esto se debe tener un conjunto de datos generados con cada distribución. El estudio se hizo con datos obtenidos de tres maneras.

1. Se simularon muestras de diferentes tamaños usando las distribuciones, Weibull(1,0; 1,2), Beta(4,4;13,3), log-normal(0,1) y PG(-0,2;1,0). (Ver programas computacionales: anexos 1 y 2 respectivamente).
2. Se utilizó un conjunto de datos reales de un estudio hidrológico facilitados por ingenieros del Ministerio de Ambiente medidos en la estación del M.A.R.N, estación Puerto la Cruz ubicada en el municipio Sotillo de la ciudad de Puerto la Cruz, Estado Anzoátegui Venezuela, cuya distribución se demuestra, usando métodos de bondad de ajuste, que es PG2 con parámetros desconocidos, los cuales se estimaron usando el método de máxima verosimilitud; se determinaron los ICP y compararon los resultados obtenidos. (Ver programa computacional en el anexo 3).
3. Por último, como el objetivo principal del trabajo fue la utilización del modelo PG2, éste fue considerado para evaluar su aplicabilidad en la estimación de los índices de capacidad de procesos; dicha estimación se hizo usando los métodos de Percentil de Clements, Percentil de Burr y el método tradicional

(suponiendo que existe normalidad en los datos simulados). Para el logro de este último procedimiento fueron considerados estimadores puntuales y los intervalos bootstrap: percentil bootstrap (PB), estándar bootstrap (EB) y percentil sesgo-correctado bootstrap (PSCB). (Ver programa computacional en el anexo 4)

4.2.2. Estimación y análisis de los Índices de Capacidad de Procesos estimado mediante diferentes métodos

4.2.2.1 Metodología de simulación

Se simularon conjuntos de 30 muestras, tamaño 100, para el conjunto de distribuciones Weibull(1,0;1,2), Beta(4,4;13,3), Log-normal(0,1) y PG2(-0,2;1,0). Se siguió la metodología presentada en el capítulo 2, (Liu-Chen (2006)). Como el límite de especificación superior, U_t , no es conocido se calculó usando la fórmula 4.1, este valor del límite de especificación superior depende de un valor fijo o target del C_{pu} , tal como se observa en la fórmula 4.1.

$$U_t = C_{pu} (\bar{X}_{0,99865} - X_{0,5}) + X_{0,5} \quad (4.1)$$

donde, $X_{0,99865}$ y $X_{0,5}$ son los puntos percentiles en cada distribución. Además, se calcularon los valores de las medias, desviaciones estándares, coeficientes de asimetrías y kurtosis en cada conjunto de muestra. Los valores fijos para C_{pu} fueron {1,0; 1,5; 2,0}.

Se utilizaron los métodos de Percentil de Clements y Burr para estimar los índices C_p y C_{pk} . Además, se calcularon los valores promedios para U_t , medias muestrales, desviaciones muestrales, percentiles, coeficientes de asimetrías, kurtosis. Se seleccionaron los parámetros c y k basados de acuerdo a los valores de asimetría y kurtosis que se encuentran tabulados en Liu y Chen (2006), para luego determinar los valores de los percentiles estandarizados 0,135, 0,5 y 99,865 mediante el método de Clements, (abreviados como $Z_{0,00135}$; $Z_{0,5}$ y $Z_{0,99865}$ respectivamente) y los percentiles de la distribución Burr, (abreviados $BZ_{0,00135}$; $BZ_{0,5}$ y $BZ_{0,99865}$).

Con los resultados antes mencionados, se calcularon los valores para L_p , M y U_p usando las fórmulas (2.11)-(2.13) en cada método.

Finalmente, se determinaron los valores promedios de los ICP con percentil de Clements (CC_p y CC_{pk}); percentil de Burr, (BC_p y BC_{pk}). Los programas que se utilizaron para generar todos estos resultados se presentan en el anexo 1 (creado con el software R) y en el anexo 2 (creado con el software S-plus). Es de hacer notar que, ambos anexos son presentados para las distribuciones Weibull(1,0; 1,2) y Beta(4,4; 13,3); pero en general, para cualquiera otra distribución que se quiera estudiar, el procedimiento a seguir es similar, solo se debe cambiar el nombre de la distribución conocida donde sea necesario. Sin embargo, para simular la muestra mediante la distribución Pareto Generalizada con 2 parámetros es diferente; dicha distribución se debe construir dentro del programa tal como se presenta al comienzo del anexo 3.

Los resultados obtenidos para cada distribución se presentan en las tablas 4.1 hasta la 4.4, respectivamente. De manera similar, se presentan en las figuras 4.5 hasta la 4.8 las gráficas de cajas para cada distribución estudiada.

Tabla 4.1: Resultados promedios para la Distribución Weibull(1,0; 1,2)

		Burr		Clements	
		BC_p		CC_p	
C_{pu}	U_t	Media	Desviación	Media	Desviación
1,0	5,7061	0,8876	0,0777	0,9058	0,08798
1,5	8,1309	1,2723	0,1183	1,2981	0,1290
2,0	10,556	1,6570	0,1597	1,6903	0,1711

		BC_{nk}		CC_{nk}	
C_{nu}	U_t	Media	Desviación	Media	Desviación
1,0	5,7061	0,8163	0,1311	0,8306	0,1281
1,5	8,1309	1,1219	0,3040	0,9718	0,2459
2,0	10,556	1,3478	0,4776	1,0057	0,3261

Tabla 4.2: Resultados promedios para la Distribución Beta(4,4; 13,3)

		Burr		Clements	
		BC _n		CC _n	
C _{mi}	U _t	Media	Desviación	Media	Desviación
1,0	5,7061	0,9137	0,1209	0,9192	0,1215
1,5	8,1309	1,2029	0,1579	1,2293	0,1714
2,0	10,556	1,3721	0,3056	1,5394	0,222

		BC _{nk}		CC _{nk}	
C _{mi}	U _t	Media	Desviación	Media	Desviación
1,0	5,7061	0,7987	0,1209	0,8019	0,1412
1,5	8,1309	1,0459	0,2949	0,9278	0,3117
2,0	10,556	1,1585	0,4347	0,9865	0,4333

www.bdigital.ula.ve

Tabla 4. 3: Resultados promedios para la Distribución Log-normal (0; 1)

		Burr		Clements	
		BC _n		CC _n	
C _{mi}	U _t	Media	Desviación	Media	Desviación
1,0	5,7061	0,9212	0,08796	0,9489	0,0862
1,5	8,1309	1,3282	0,13094	1,3679	0,1261
2,0	10,556	1,7351	0,1744	1,7244	0,2801

		BC _{nk}		CC _{nk}	
C _{mi}	U _t	Media	Desviación	Media	Desviación
1,0	5,7061	0,8672	0,0827	0,9112	0,0804
1,5	8,1309	1,2866	0,1217	1,1838	0,1996
2,0	10,556	1,6712	0,2134	1,3643	0,3944

c.c Reconocimiento

Tabla 4.4: Resultados promedios para la Distribución PG2 (-0,2; 1,0)

C_{mu}	U_t	Burr		Clements	
		Media	Desviación	Media	Desviación
1,0	5,7061	0,8031	0,0853	0,8149	0,0993
1,5	8,1309	1,1898	0,12695	1,2070	0,1469
2,0	10,556	1,5764	0,1689	1,5366	0,3014

C_{mu}	U_t	BC_{pk}		CC_{pk}	
		Media	Desviación	Media	Desviación
1,0	5,7061	0,6735	0,1727	0,6423	0,1909
1,5	8,1309	0,7409	0,2930	0,7021	0,3177
2,0	10,556	0,7849	0,4029	0,7405	0,4214

www.bdigital.ula.ve

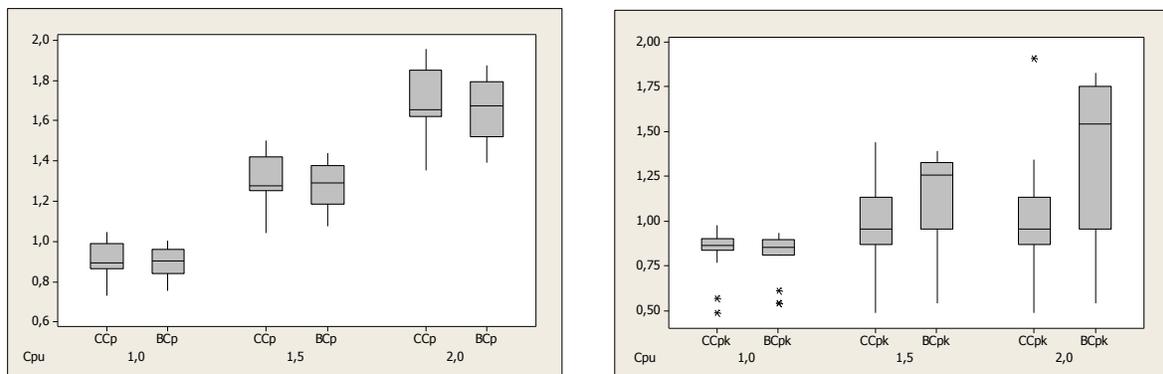


Figura 4.5: Diagrama de cajas para los índices (CC_p ; BC_p) y (CC_{pk} ; BC_{pk}) respectivamente para una Distribución Weibull(1,0;1,2)

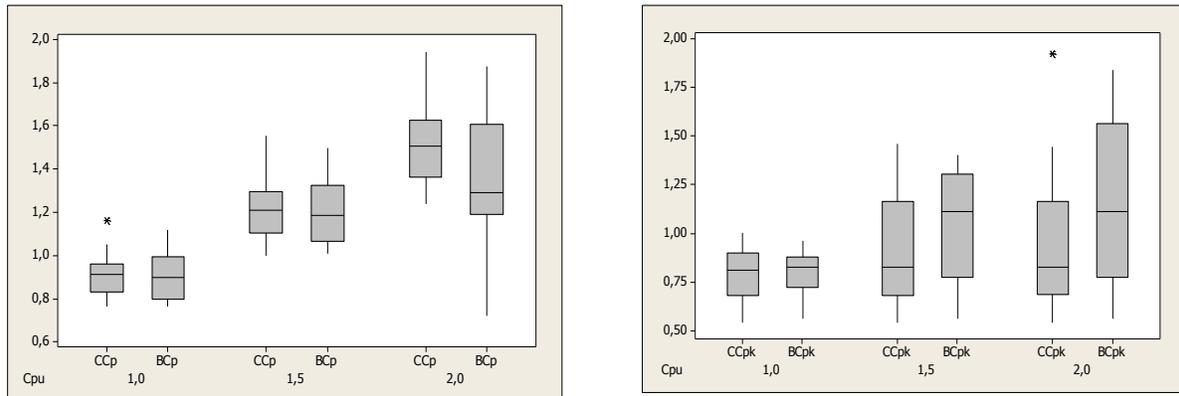


Figura 4.6: Diagrama de cajas de los índices (CC_p ; BC_p) y (CC_{pk} ; BC_{pk}) respectivamente para la Distribución Beta(4,4;13,3)

www.bdigital.ula.ve

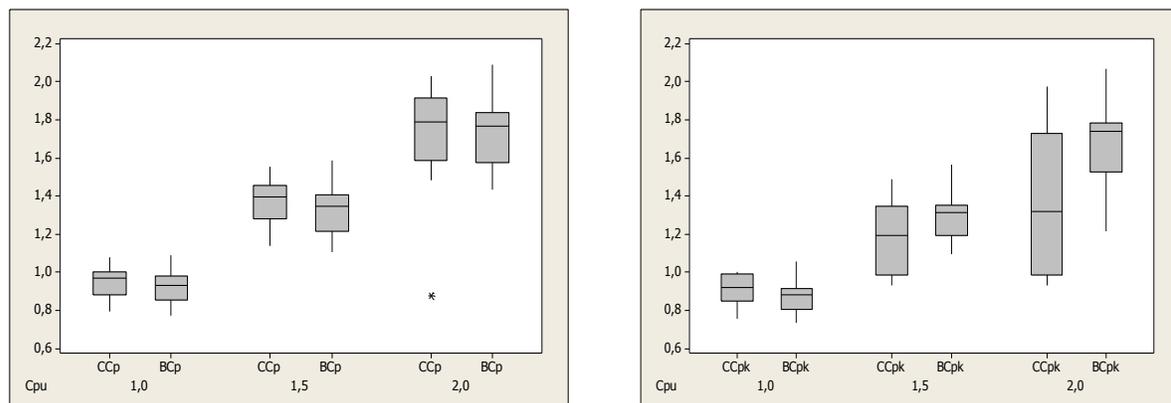


Figura 4.7: Diagrama de cajas de los índices (CC_p ; BC_p) y (CC_{pk} ; BC_{pk}) respectivamente para la Distribución Log-normal(0,1)

c.c Reconocimiento

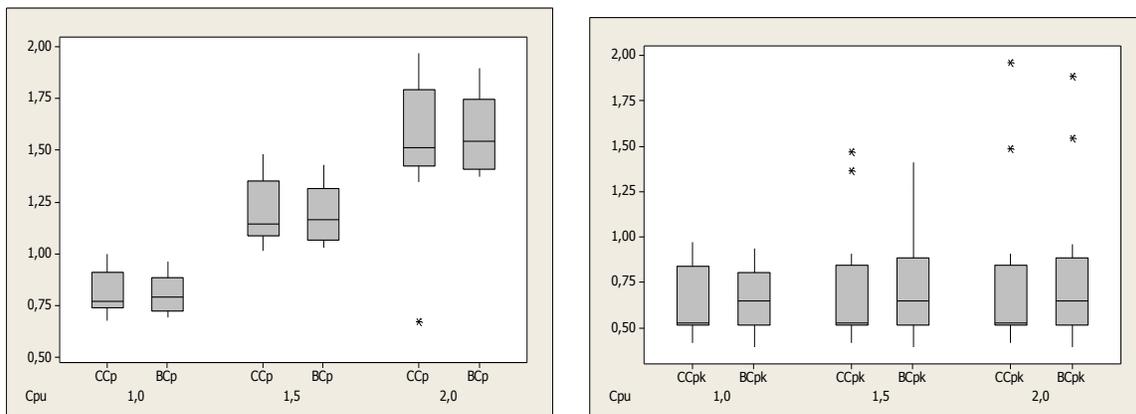


Figura 4.8: Diagrama de cajas de los índices $(CC_p; BC_p)$ y $(CC_{pk}; BC_{pk})$ respectivamente para la DPG2 $(-0,2;1,0)$

- Comparación de Resultados

Para considerar el desempeño de los índices de capacidad de procesos obtenidos mediante los métodos de percentiles de Clements (CC_p , CC_{pk}) y percentiles de Burr (BC_p y BC_{pk}) respectivamente, se tomó en cuenta el efecto ocasionado por la desviación estándar en cada distribución; de acuerdo a los resultados presentados en las tablas 4.1 hasta la 4.4; se compararon los índices CC_p y BC_p análogamente se compararon CC_{pk} y BC_{pk} .

- Respecto a la media y a los valores targets del C_{pu} . En general, el índice de Clements, CC_p , produce mejores resultados cuando la distribución considerada presenta desviación severa de la normalidad; esto se refleja en los valores de los ICP más próximos a los valores del C_{pu} . Es decir, según el método de Clements, el modelo PG produce resultados más próximos al C_{pu} , seguido del modelo Weibull, luego Beta y finalmente Log-normal; esto indica que posiblemente mientras más asimétrica es la distribución de los datos, se esperan resultados de CC_p más próximos a los valores targets. Lo que lleva a pensar que su estimación es mejor en comparación con los resultados obtenidos con el método de percentil de Burr. Por otro lado, las distribuciones con asimetrías moderadas generan mayor dispersión en

los datos; de acuerdo a lo obtenido en las simulaciones un orden dado en cuanto a las distribuciones que estiman los ICP de menor a mayor confiabilidad es el siguiente Log-normal, Beta, Weibull y PG2.

- b. Similarmente, tomando en cuenta la precisión promedio, el índice estimado por el método de Bur, BC_{pk} estima mejor que el CC_{pk} ; es decir, la precisión en este tipo de índice es mayor considerando el método de percentiles de Burr que en el Clements. Finalmente, la $DPG2(-0,2; 1,0)$ produce desviación más pequeña considerando el método de Burr en contraparte con la distribución Log-normal las cuales son un poco mayor que las demás. En los diagramas de cajas se recogen resultados como la mediana, primer y tercer cuartil, aparte de los outliers observados en cada distribución estudiada.

4.2.2.2 Caso Real

En esta sección se ilustran ambos métodos en el ajuste de datos muestrales a una distribución no normal desconocida. La distribución obtenida se utiliza para estimar los índices de capacidad de procesos.

Se consideró una serie de datos observacionales sobre intensidad máxima de precipitación diaria (mm/horas) medidas en la estación de M.A.R.N, estación Puerto la Cruz ubicada en el municipio Sotillo de la ciudad de Puerto la Cruz, Estado Anzoátegui, Venezuela. Dicha información refleja las precipitaciones caídas en el Estado con un tiempo máximo de duración de 1 hora, ocurridas durante los años 1961 hasta 2000. La serie facilitada por ingenieros de la institución, contiene 198 observaciones. El objetivo es evaluar un proceso hidrológico ocurrido en el período de tiempo 1961-2000. La aleatoriedad de los datos se verifica para asegurar la estabilidad del proceso. Los datos muestrales se ordenan de manera ascendente. La distribución de frecuencia y el resumen se presentan en la tabla 4.5. En la figura 4.9, se representa el histograma de la serie de datos.

- Aplicación de la distribución no normal

Para verificar la bondad de ajuste de la distribución en estudio que resultó ser una $DPG2$, se siguieron los siguientes pasos.

Paso 1: Los valores de los parámetros β y α son desconocidos, usando el software S-plus, se estimaron mediante el método máxima verosimilitud; los resultados fueron los siguientes, $\hat{\beta} = 20,655$ y $\hat{\alpha} = 0,037$.

Paso 2: Se llevó a cabo los tests de Cramer Von Mises (W^2) y Anderson Darling (A^2), $W^2=1,3676$ ($p < 0,005$) y $A^2=8,664$ ($p < 0,005$).

De la curva de la distribución Pareto ajustada, se hallaron los valores de los percentiles 0,00135; 0,5 y 0,99865, respectivamente. Una vez obtenidos estos, se determinaron los coeficientes de asimetría y kurtosis, para luego estimar los percentiles estandarizados, $L_p=1,00$; $M=16,00$ y $U_p=105,341$. Algunos resultados se recogen en la tabla 4.5. Los límites de especificaciones inferior y superior considerados para este estudio fueron tomados como, $L_t=1$ y $U_t=107$.

Paso 3: Observando los valores de los p-value se puede inferir que dicha serie de datos es ajustada por una distribución Pareto con parámetros β y α .

Paso 4: Finalmente, se estimaron los índices de capacidad normales y los no normales utilizando las fórmulas (2.2)-(2.5) y (2.11)-(2.15), respectivamente. Los resultados se muestran en la tabla 4.6.

De acuerdo a los datos obtenidos para los índices respectivos, se puede observar que la serie de datos proporcionan índices de capacidad para un proceso estadísticamente no capaz, ya que estos son menores a los valores estándares. Además, los índices obtenidos con el método de percentil de Clements ($\hat{C}\hat{C}_p$ y $\hat{C}\hat{C}_{pk}$) y los obtenidos con el método de Burr ($\hat{B}\hat{C}_p$ y $\hat{B}\hat{C}_{pk}$), son mayores que los tradicionales (\hat{C}_p y \hat{C}_{pk}). Por otro lado, el método de Burr genera mejores resultados cuando la distribución presenta desviación severa de la normalidad, esto permite asegurar que la asimetría en los datos observacionales es influyente a la hora de determinar dichos índices. Se aprecia la diferencia entre considerar una distribución normal de los datos y tomar una distribución PG. El programa construido para desarrollar este estudio y obtener la información antes dada es el anexo 3 que se encuentra en la sección de anexos al final.

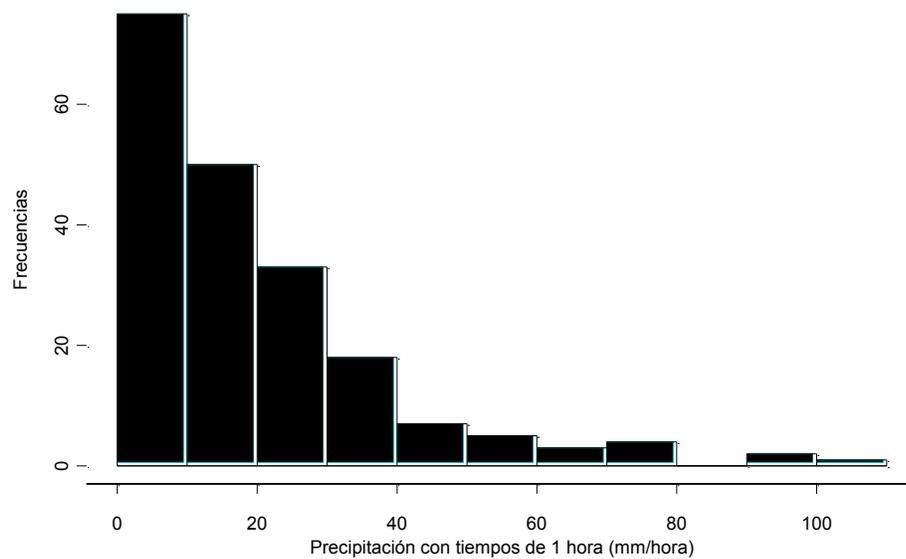


Figura 4.9: Histograma de la serie de datos con tiempo de 1 hora máxima de precipitación.

Fuente: M.A.R.N., estación Puerto la Cruz. Municipio Sotillo

Tabla 4.5: Resumen estadístico para una serie de datos obtenida de un proceso hidrológico con un tiempo de duración de 1 hora de lluvia.

Min	Q_1	Med	\bar{x}	Q_3	Max	s	Asim	Kurtosis
1,00	5,25	16,00	19,92	26,00	108,00	19,23141	1,798	4,50
L_p	U_p	α_3	α_4	L_t	U_t	$\hat{\beta}$	$\hat{\alpha}$	
1,00	105,341	1,87	7,3	1,00	107,00	20,655	0,037	

Tabla 4.6: Valores de los índices obtenidos con los métodos de Percentil de Clements, de Burr y los tradicionales

Índices de Capacidad usando Percentil de Clements.	Índices de Capacidad usando Percentil de Burr	Índices de Capacidad de procesos normales
$\hat{C}_p = 0,955$	$\hat{B}C_p = 0,997$	$\hat{C}_p = 0,919$
$\hat{C}_{pk} = 0,911$	$\hat{C}C_{pk} = 0,935$	$\hat{C}_{pk} = 0,328$

4.2.3 Intervalos bootstrap para estimar ICP usando la DPG con 2 parámetros.

En esta sección se trabaja con el modelo de la Distribución Pareto Generalizada con 2 parámetros (DPG2), para evaluar el desempeño de los índices de capacidad de procesos no normales. Se tomó en cuenta el método de Percentiles de Clements, el percentiles de Burr y el método tradicional. Se consideraron estimadores puntuales y por intervalos bootstrap (percentil bootstrap (PB), estándar bootstrap (EB) y percentil sesgo-correctado bootstrap (PSCB)), generados mediante simulaciones y considerando una muestra cuya distribución es PG2. En Valdiviezo y Fermín (2010), se resume este procedimiento. Allí, se simulan intervalos de 95% y se analiza la capacidad de cubrimiento que tienen estos intervalos. Los métodos no normales considerados son Percentil de Burr, de Clements y Normal. Además se tomó una muestra real de tamaño 198 proveniente de una distribución Pareto Generalizada.

A continuación se describen los intervalos bootstrap, los cuales se utilizan en la evaluación del desempeño de los ICP bajo distribución no normal.

- **Métodos Bootstrap**

Franklin y Wasserman (1998), desarrollaron métodos de bootstrap para determinar intervalos de confianza considerando distribuciones normales y no normales. Efron y Tibshirani (1986), consideraron estos tres tipos de intervalos de confianza bootstrap, los cuales se describen más adelante.

Para construir los intervalos de confianza bootstrap, se consideraron muestras de tamaño n , (x_1, x_2, \dots, x_n) , tomada de un proceso y se determinan los índices de capacidad, ya sean por el método tradicional (C_{pk}), el de Burr, (BC_{pk}), o el de Clements, (CC_{pk}). En función a esto, se generó una muestra bootstrap (B) de tamaño 1000, sin reemplazamiento para cada uno de los diferentes índices y luego se ordena de mayor a menor, lo cual constituye la distribución de los índices respectivos.

A continuación se definen los tres diferentes intervalos bootstrap utilizados.

a. Estándar Bootstrap (EB)

Sea $\hat{C}^*(i)$, el i -ésimo estimador bootstrap ordenado del índice C^* con $i=1,2,\dots,B$. y sea,

$$\bar{c}^* = \frac{1}{B} \sum_{i=1}^B \hat{c}^*(i)$$

$$y \quad S_c^* = \left[\frac{\sum_{i=1}^B (\hat{c}^*(i) - \bar{c}^*)^2}{B-1} \right]^{\frac{1}{2}}$$

Con la cantidad S_c^* es un estimador de la desviación estándar de \hat{C} , entonces (si la distribución de \hat{C} es aproximadamente normal) el $(1-2\alpha)100\%$, intervalo de confianza bootstrap estándar está dado por,

$$\bar{c}^* \pm z_\alpha S_c^*, \quad (4.2)$$

donde, z_α es el $(1-2\alpha)100\%$, valor del percentil de la distribución normal estándar, en nuestro caso 1,96.

b. Percentil Bootstrap (PB)

Para la colección ordenada considerada, el estimado $\hat{c}^*(i)$ del $(1-2\alpha)100\%$ intervalo de confianza bootstrap percentil, está dado por.

$$[\hat{c}^*(\alpha B), \hat{c}^*((1-\alpha)B)] \quad (4.3)$$

a. Percentil Sesgo-Corregido Bootstrap (PSCB)

Para este método puede darse el caso que la distribución bootstrap obtenida para las B muestras pueda ser sesgada, con este método se trata de corregir este sesgo.

Primeramente, usando la distribución ordenada de \hat{C}^* , se calcula la probabilidad siguiente,

$$p_0 = \Pr[\hat{C}^* \leq \hat{c}] \quad (4.4)$$

donde, \hat{c} está dada por el valor del índice de capacidad estimado de la muestra original de donde se tomó la muestra bootstrap. Se calculan los siguientes valores.

$$z_0 = \Phi^{-1}(p_0) \quad (4.5)$$

$$P_{LEI} = \Phi(2z_0 - z_\alpha) \quad (4.6)$$

$$P_{LES} = \Phi(2z_0 + z_\alpha) \quad (4.7)$$

donde $\Phi(\cdot)$ es la función de distribución acumulativa para la distribución normal. Y la función $\Phi^{-1}(\cdot)$ es la inversa. Luego, el intervalo de confianza BCPB está dado por la siguiente expresión:

$$[\hat{C}^*(P_{LEI}B), \hat{C}^*(P_{LES}B)] \quad (4.8)$$

- **Estudio de simulación**

Un estudio con diversas simulaciones se realizó para la obtención de intervalos de confianzas bootstrap de 95%, con el objetivo de estimar los índices de capacidad de procesos con los métodos de Burr, Clements y Normal.

En la figura 4.10, se presenta el histograma resultante de una muestra de tamaño 198 proveniente de una distribución Pareto Generalizada con parámetros $\hat{\beta} = 20,66$ y $\hat{\alpha} = 0,44$. Así como en la tabla 4.7, se presenta un resumen estadístico de dicha muestra.

Tabla 4.7: Resumen estadístico de la muestra de observaciones provenientes de la DPG2(20,66; 0,44)

Min	Q ₁	Med	\bar{x}	Q ₃	Max	s	Asi	Kurt	$\hat{\beta}$	$\hat{\alpha}$
1,00	5,25	16,00	19,9	26,	108,0	19,23	1,89	4,47	20,6	0,44
Q _{0,001}	Q _{0,99}	α_3	α_4							
1,0	105,	1,87	7,33							

- **Simulación de los datos**

Una vez obtenidos los resultados de la tabla 4.7, se simularon 1000 muestras bootstrap sin reemplazamiento obtenidas de la DPG(20,66;0,44), para determinar luego los valores estimados de los índices de capacidad de procesos. Se consideraron tres diferentes valores

targets para un C_{pu} los cuales son $\{0,5; 1,0 \text{ y } 1,5\}$. Se tomó, L_t como el valor mínimo de la muestra simulada. Con cada valor target se obtiene el valor de U_t y los índices respectivos, los cuales se presentan en la tabla 4.8.

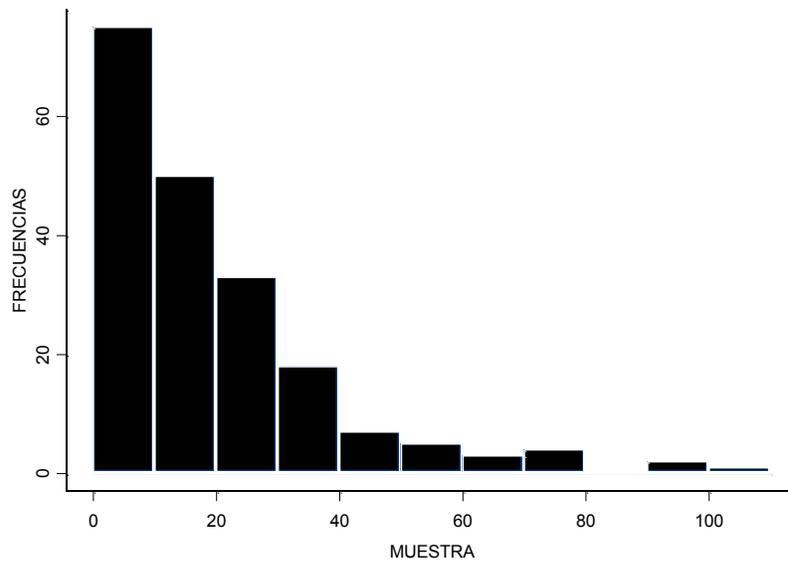


Figura 4.10: Histograma de la muestra distribuida $PG2(20,66; 0,44)$

Tabla 4.8: Valores para diferentes C_{pu} , U_t y los índices de capacidad estimados

	$C_{pu}= 0,50$	$C_{pu}= 1,0$	$C_{pu}= 1,5$
U_t	60,67	105,34	150,01
\widehat{BC}_{pk}	0,4885	0,9192	1,3499
\widehat{CC}_{pk}	0,4669	0,8949	1,3229
\widehat{C}_{pk}	0,3280	0,3280	0,3280

Se construyeron los intervalos de confianza Percentil, Estándar y el Percentil Sesgo-correctido con un nivel de confianza de 95%.

Una vez obtenidos los intervalos bootstrap, se simularon nuevamente 500 muestras por cada intervalo bootstrap y en cada caso se determinan los valores para: el porcentaje de cubrimiento, la longitud promedio y desviación estándar de dichas longitudes promedios.

Los intervalos bootstrap generados para los valores estimados de los \widehat{BC}_{pk} , \widehat{CC}_{pk} y \widehat{C}_{pk} en cada caso, se presentan en las tablas 4.9 - 4.11. Es de hacer notar, que dichos intervalos, obtenidos mediante el método tradicional utilizan, para la estimación del índice C_{pk} , a la media muestral, esto significa que su resultado no varía aun cuando se intercambien los valores para C_{pu} y U_t respectivamente.

Tabla 4.9: Proporción del cubrimiento del 95% para el \widehat{BC}_{pk}

$\widehat{BC}_{pk}=0,4885$	Cubrimiento	Ancho	Desviación
Percentil: [0,4092;0,5992]	0,97	0,505	0,0378
Estándar: [0,4005;0,5876]	0,94	0,493	0,0366
Percentil Corregido: [0,4068;0,5927]	0,93	0,499	0,0367
$\widehat{BC}_{pk}=0,9192$			
Percentil: [0,7809;1,1308]	0,99	0,9584	0,0691
Estándar: [0,7561;1,1062]	0,97	0,9328	0,0691
Percentil Corregido: [0,7604;1,1167]	0,97	0,9409	0,0715
$\widehat{BC}_{pk}=1,3499$			
Percentil: [1,1453;1,6680]	0,94	1,4033	0,1035
Estándar: [1,1204;1,6184]	0,97	1,3659	0,0978
Percentil Corregido:	0,97	1,3837	0,0972

Tabla 4.10: Proporción del cubrimiento para el 95% para el \widehat{CC}_{pk}

$\widehat{CC}_{pk}=0,4669$	Cubrimiento	Ancho	Desviación
Percentil: [0,3899;0,5725]	0,95	0,4794	0,0364
Estándar: [0,3824;0,5633]	0,98	0,4722	0,0352
Percentil Corregido:	0,93	0,4643	0,0352
$\widehat{CC}_{pk}=0,8949$			
Percentil: [0,7542;1,0981]	0,92	0,9255	0,0673
Estándar: [0,7375;1,0819]	0,92	0,9107	0,0674
Percentil Corregido:	0,97	0,9112	0,0665
$\widehat{CC}_{pk}=1,3229$			
Percentil: [1,1255;1,6274]	0,91	1,3788	0,0986
Estándar: [1,0934;1,6019]	0,96	1,3463	0,1005
Percentil Corregido: [1,1036;1,5871]	0,94	1,3465	0,0958

Tabla 4.11: Proporción del cubrimiento para el 95% para el \hat{C}_{pk}

$\hat{C}_{pk} = 0,3280$	Cubrimiento	Ancho	Desviación
Percentil: [0,294; 0,3768]	0,87	0,3348	0,0165
Estándar: [0,29; 0,3739]	0,85	0,3321	0,0165
Percentil Corregido: [0,2884; 0,3689]	0,83	0,3287	0,0158

A continuación se presentan las figuras 4.11, 4.12 y 4.13 correspondientes al cubrimiento de cada intervalo obtenido en la tabla 4.11.

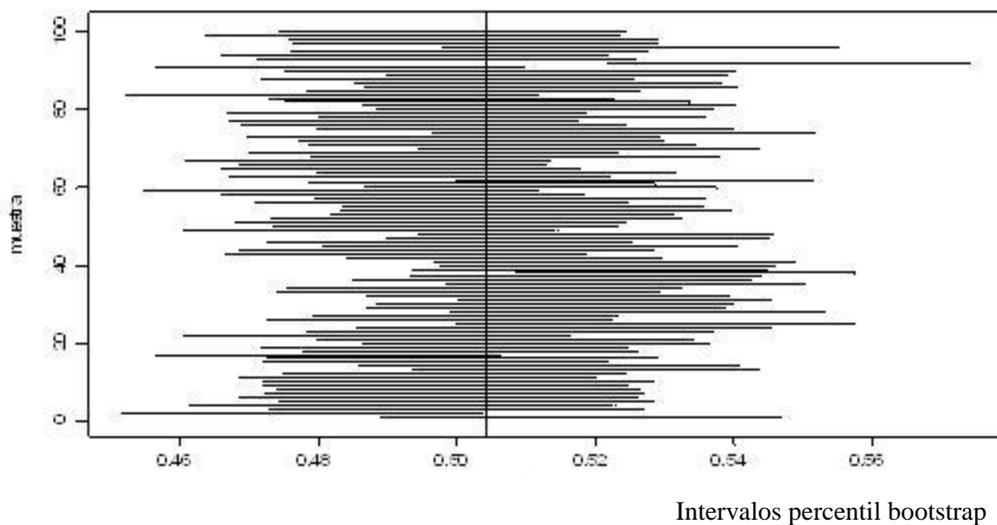


Figura 4.11: Cubrimiento de los intervalos Percentil Bootstrap

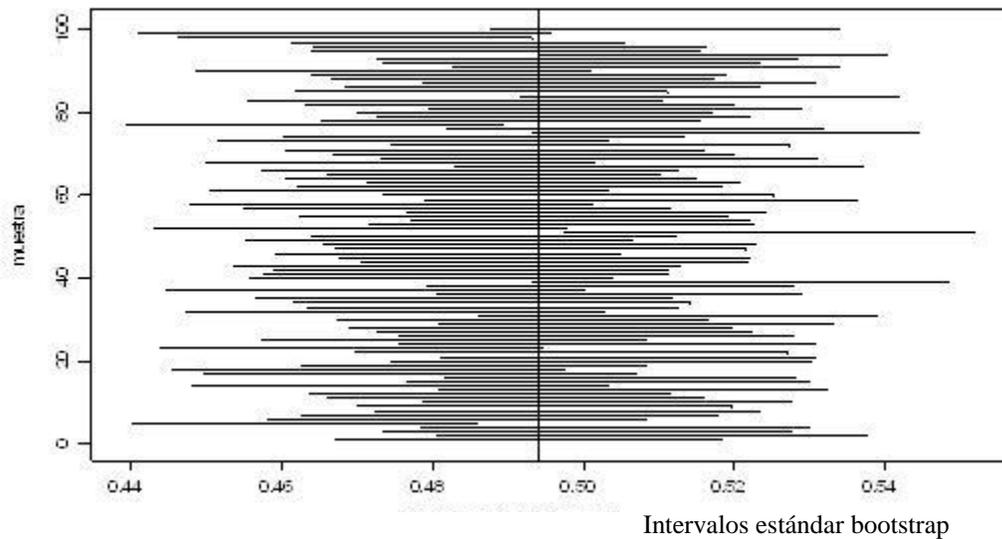


Figura 4.12: Cubrimiento de los intervalos Estándar Bootstrap

www.bdigital.ula.ve

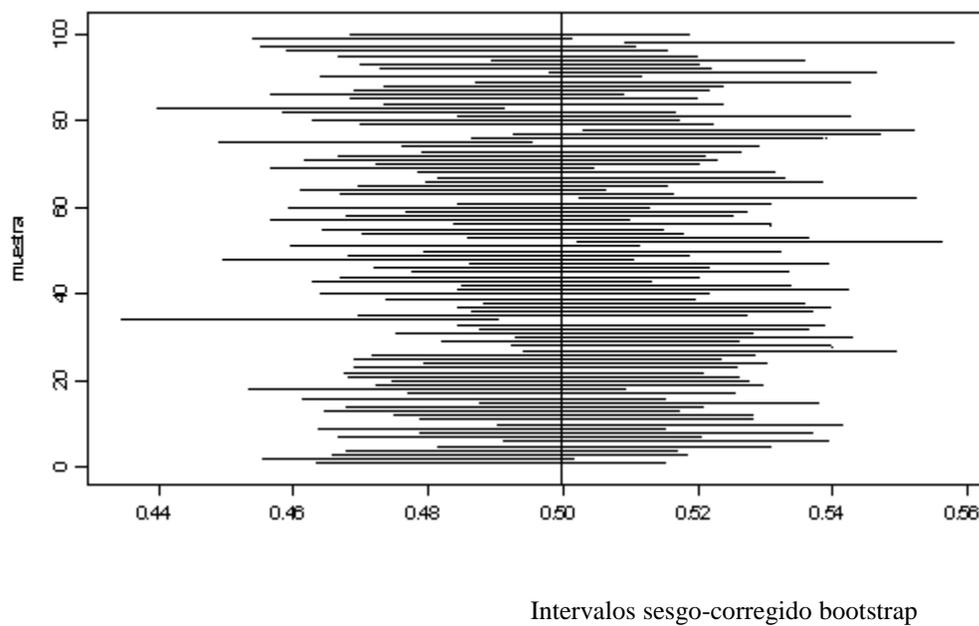


Figura 4.13: Cubrimiento de los intervalos Percentil Sesgo-Corregido Bootstrap

c.c Reconocimiento

Discusión de resultados

Para las muestras simuladas con la DPG(20,66; 0,44) se lograron estimar los intervalos bootstrap. El procedimiento se repitió de manera sistemática tomando diferentes valores targets para C_{pu} , {0,5; 1,0; 1,5}.

- a. Para $C_{pu}=0,5$ se lograron los siguientes resultados; $\hat{B}C_{pk}=0,4885$, $\hat{C}C_{pk}=0,4669$ y $\hat{C}_{pk}=0,3280$. De acuerdo a lo observado en las tablas (4.8) - (4.10), el intervalo PB para el índice $\hat{B}C_{pk}$ presenta mayor porcentaje de cubrimiento y mayor ancho promedio, mientras que el estimado para el índice \hat{C}_{pk} tiene menor porcentaje de cubrimiento, en cuanto a la desviación promedio, es mayor en los índices no normales que en el índice normal. En el intervalo EB, el porcentaje de cubrimiento resulta mayor en el $\hat{C}C_{pk}$ que en los otros dos índices, así como su ancho promedio; las desviaciones son relativamente iguales entre los índices no normales, pero significativamente alta en comparación con el índice \hat{C}_{pk} estimado. Para el intervalo PSCB se mantienen los porcentajes de cubrimientos relativamente iguales entre los índices no normales, pero aún más alto en comparación con el índice \hat{C}_{pk} estimado; algo similar ocurre en la desviación promedio, la cual es más alta en los índices no normales. En general el PB es mejor estimador, en promedio, según los métodos de percentiles de Burr y de Clements, es más amplio que los intervalos EB y PSCB.
- b. Por otro lado, para $C_{pu}=1,0$ se lograron los resultados $\hat{C}C_{pk}=0,9192$; $\hat{B}C_{pk}=0,8949$ y $\hat{C}_{pk}=0,3280$. El intervalo EB continúa teniendo mayor porcentaje en cubrimiento de acuerdo al índice estimado $\hat{B}C_{pk}$, similarmente su ancho promedio. Las desviaciones promedios son similares entre los estimados no normales mientras que el estimado bajo suposición normal se mantiene igual en todo los casos.
- c. Finalmente, para $C_{pu}=1,5$ se lograron los resultados $\hat{B}C_{pk}=1,3499$; $\hat{C}C_{pk}=1,322$ y $\hat{C}_{pk}=0,3280$, observándose que sigue siendo el índice estimado $\hat{B}C_{pk}$ el de mayor porcentaje en cubrimiento y el de mayor ancho promedio. Las desviaciones se mantienen muy similares entre los índices no normales, pero mayor en

comparación con el índice normal. Para los diferentes valores targets fijados para $C_{pu}(0,50;1,0; 1,5)$, el intervalo PSCB es más confiable, en promedio, según los métodos de Percentiles de Burr y de Clements.

En general, se ha descrito el procedimiento que involucra un método de estimación de ICP no normales, como lo es el método de Clements modificado, usando la distribución Pareto Generalizada con 2 parámetros; esto con el propósito de evaluar su desempeño, además se simuló conjuntos de datos considerando distribuciones con asimetrías leves, moderadas y severas para luego ser comparados entre sí, resultando que el método de percentil de Burr es ligeramente mejor que el método de Clements en distintos casos. En general, el método de Burr arroja resultados con dispersión menor en las distribuciones con asimetría severa de su normalidad en comparación con el método de Clements.

En cuanto a la precisión, las distribuciones con asimetrías moderadas representan mejores modelos para estimar los ICP, sobre todo cuando el valor del parámetro de forma se aproxima a 0,50 en contra parte de cuando la asimetría de la distribución es severa y los valores del parámetro de forma estimado se aproximan a 1. Por otro lado, para el conjunto de datos reales se ajustó la DPG2, mediante test de bondad se verificó el ajuste del modelo, se determinó los valores de los parámetros por el método de máxima verosimilitud y finalmente se calcularon los ICP, cuando no son normales y suponiendo que sí lo son; observándose en la tabla 4.6, que los índices obtenidos con el método de percentil de Clements y los obtenidos con el método de Burr, arrojan mejores resultados, al menos más próximos a los valores target fijados de C_{pu} , mientras que para los índices calculados con el método tradicional $(\hat{C}_p, \hat{C}_{pk})$, los resultados son más pequeños. Esto es indicativo de que a la hora de decidir analizar nuevamente un proceso usando los ICP, es necesario chequear la condición de normalidad antes de decidir cualquier cosa, ya que dicho proceso puede ocasionar pérdidas, bien sea porque se continúe o se detenga dependiendo de la situación.

4.3 ÍNDICES DE CAPACIDAD ESTIMADOS MEDIANTE LA FUNCION DE DENSIDAD DE LA DPG2

En esta sección del trabajo se presenta otra forma de estimar los índices C_p y C_{pk} utilizando el modelo de la distribución Pareto Generalizada con 2 parámetros. Consiste en introducir en las expresiones 2.2 hasta la 2.5, la fórmula correspondiente de la función de distribución PG2 y obtener así otra expresión para C_p y C_{pk} . Se realiza el estudio mediante simulación de datos y los métodos para estimarlos son, el método de Percentil de Clements, Percentiles de Burr y el método exacto. Otras distribuciones que se tomaron en cuenta para la comparación fueron las distribuciones Weibull, Log-normal y Beta.

- Aplicación del modelo de la DPG

Wu et al. (1998) llevaron a cabo una investigación, donde indica que el método de estimación de ICP no puede generar medidas confiables cuando la distribución desconocida de los datos es asimétrica. Se propone a continuación la estimación de ICP mediante la distribución Pareto Generalizada, usando la fórmula de la función de distribución para estimar los valores de los cuantiles y las fórmulas (2.2)- (2.5) para la estimación de ICP.

De la fórmula (2.2) se tiene,

$$C_p = \frac{U_t - L_t}{x(F_{99,865}) - x(F_{0,135})} = \begin{cases} \frac{\alpha\beta^{-1}[U_t - L_t]}{(1 - F_{0,135})^\alpha - (1 - F_{99,865})^\alpha} & \text{para } \alpha \neq 0 \\ \frac{[U_t - L_t]}{\beta \left[\log \left(\frac{(1 - F_{0,135})}{(1 - F_{99,865})} \right) \right]} & \text{para } \alpha = 0 \end{cases} \quad (4.9)$$

Similarmente de las fórmulas (2.2) y (2.3), se tiene que,

$$C_{pu} = \frac{U_t - M}{U_p - M} = \begin{cases} \frac{U_t - \frac{\beta}{\alpha} [1 - (1 - F_{50})^\alpha]}{\frac{\beta}{\alpha} [(1 - F_{50})^\alpha - (1 - F_{99,865})^\alpha]} & \text{para } \alpha \neq 0 \\ \frac{U_t + \beta \log(1 - F_{50})}{\beta \left[\log \left(\frac{(1 - F_{50})}{(1 - F_{99,865})} \right) \right]} & \text{para } \alpha = 0 \end{cases} \quad (4.10)$$

$$C_{pl} = \frac{M - L_t}{M - L_p} = \begin{cases} \frac{\frac{\beta}{\alpha} [1 - (1 - F_{50})^\alpha] - L_t}{\frac{\beta}{\alpha} [(1 - F_{50})^\alpha - (1 - F_{99,865})^\alpha]} & \text{para } \alpha \neq 0 \\ \frac{-\beta \log(1 - F_{50}) - L_t}{\beta \left[\log \left(\frac{(1 - F_{0,135})}{(1 - F_{50})} \right) \right]} & \text{para } \alpha = 0 \end{cases} \quad (4.11)$$

Luego, se tiene de (2.4),

$$C_{pk} = \text{Min}(C_{pu}, C_{pl})$$

4.4 ÍNDICES DE CAPACIDAD DE PROCESOS ESTIMADOS MEDIANTE EL MÉTODO EXACTO

Este método permite también evaluar el desempeño de un proceso mediante los ICP. Czarski (2008), presenta una relación entre los percentiles $X_{0,00135}$, $X_{0,5}$ y $X_{0,99865}$, las cuales se definen a continuación.

$$\int_{-\infty}^{X_{0,00135}} f(x) dx = 0,00135 \quad (4.12)$$

$$\int_{-\infty}^{X_{0,50}} f(x) dx = 0,50 \quad (4.13)$$

$$\int_{-\infty}^{X_{0,99865}} f(x) dx = 0,99865 \quad (4.14)$$

Para determinar los valores de los percentiles usando las fórmulas (4.12)-(4.14), se requiere el conocimiento de la función de densidad de la distribución. De manera general,

$$\int_{-\infty}^{x_i} f(x) dx = i \text{ con } i = \{0,00135; 0,50; 0,99865\} \text{ y además, } \int_{-\infty}^{x_i} f(x) dx := F(x)$$

4.4.1 Índices de capacidad basados en el método exacto considerando la DPG2

La función de densidad de la distribución PG2 está definida como, $F(x) = 1 - \left(1 - \alpha \frac{x}{\beta}\right)^\alpha$.

Así, los valores de los percentiles, según el método exacto y usando la DPG2, están dados de la siguiente manera.

$$\frac{\beta}{\alpha} \left[1 - (1 - 0,00135)^\alpha \right] = X_{0,00135} \quad (4.15)$$

$$\frac{\beta}{\alpha} \left[1 - (1 - 0,50)^\alpha \right] = X_{0,50} \quad (4.16)$$

$$\frac{\beta}{\alpha} \left[1 - (1 - 0,99865)^\alpha \right] = X_{0,99865} \quad (4.17)$$

4.4.2 Análisis de simulaciones

En esta sección se realizaron una serie de simulaciones conducidas con muestras de tamaño $n=25, 50, \text{ y } 100$ para una familia de distribuciones Pareto Generalizada con 2 parámetros; se fijó un valor para el parámetro $\beta=1$ y se tomaron valores diferentes para $\alpha>0,5$; estos se denotaron $PG2(\alpha_n, 1)$. Los ICP estimados con este modelo son denotados respectivamente PC_P y PC_{pk} ; mientras que, los obtenidos con el método de percentil de Burr se denotaron, BC_P y BC_{pk} , los estimados con el método de Percentil de Clements, CC_P y CC_{pk} , finalmente, los estimados mediante el método exacto se denotaron, EC_P y EC_{pk} . Siguiendo la metodología presentada por Liu-Chen (2006), se consideraron valores fijos para C_{pu} $\{1,0; 1,50; 2,0\}$. En cada corrida se determinaron los valores muestrales de \bar{x} , s , percentil inferior (L_t), percentil superior (U_t) y mediana (Med).

En las tabla 4.12, 4.13 y 4.14, se muestran resultados de los ICP estimados con el modelo PG2 para diferentes valores del parámetro α , muestras distintas ($n=25, 50, 100$) y $\beta=1$.

Tabla 4.12: Resultados de los ICP con una muestra simulada ($n=25$) mediante el modelo $PG2(\alpha_n, 1)$

α_n	$C_{pu}=1,00$	$C_{pu}=1,50$	$C_{pu}=2,00$
$PC_p(0,52;1)$	0,4800	0,6459	0,8117
$BC_p(0,52;1)$	0,8839	1,1893	1,4947
$CC_p(0,52;1)$	0,9588	1,2901	1,6214
$EC_p(0,52;1)$	0,9022	1,2138	1,5255
$PC_{pk}(0,52;1)$	0,3690	0,5213	0,5213
$BC_{pk}(0,52;1)$	0,8054	1,0946	1,0946
$CC_{pk}(0,52;1)$	0,9316	1,0173	1,0173
$EC_{pk}(0,52;1)$	0,8676	0,9782	0,9782
$PC_p(0,60;1)$	0,5425	0,7227	0,9029
$BC_p(0,60;1)$	0,8441	1,1245	1,4048
$CC_p(0,60;1)$	0,9099	1,2122	1,5145
$EC_p(0,60;1)$	1,0270	1,3819	1,7367
$PC_{pk}(0,60;1)$	0,4216	0,6016	0,6016
$BC_{pk}(0,60;1)$	0,7842	0,7842	0,7842
$CC_{pk}(0,60;1)$	0,7375	0,7375	0,7375
$EC_{pk}(0,60;1)$	0,9776	0,9776	0,9776
$PC_p(0,68;1)$	0,6161	0,8125	1,0089
$BC_p(0,68;1)$	0,8441	1,1245	1,4048
$CC_p(0,68;1)$	0,9099	1,2122	1,5145
$EC_p(0,68;1)$	1,1549	1,5539	1,9529
$PC_{pk}(0,68;1)$	0,4891	0,6819	0,6819
$BC_{pk}(0,68;1)$	0,7842	0,7842	0,7843
$CC_{pk}(0,68;1)$	0,7842	0,7842	0,7842

Tabla 4.12: Resultados de los ICP con una muestra simulada ($n=25$) mediante el modelo $PG2(\alpha_n, 1)$ (Continuación)

$EC_{nk}(0,68;1)$	0,9771	0,9771	0,9771
$PC_n(0,76;1)$	0,7027	0,9175	1,1323
$BC_n(0,76;1)$	0,8441	1,1245	1,4048
$CC_n(0,76;1)$	0,9099	1,2122	1,5145
$EC_n(0,76;1)$	1,2849	1,7289	2,1728
$PC_{nk}(0,76;1)$	0,5779	0,7622	0,7622
$BC_{nk}(0,76;1)$	0,7842	0,7842	0,7842
$CC_{nk}(0,76;1)$	0,7842	0,7842	0,7842
$EC_{nk}(0,76;1)$	0,9765	0,9765	0,9765
$PC_n(0,84;1)$	0,7973	1,0359	1,2744
$BC_n(0,84;1)$	0,8441	1,1245	1,4048
$CC_n(0,84;1)$	0,9099	1,2120	1,5145
$EC_n(0,84;1)$	1,4165	1,9059	2,3953
$PC_{nk}(0,84;1)$	0,6942	0,8425	0,8425
$BC_{nk}(0,84;1)$	0,7842	0,7842	0,7842
$CC_{nk}(0,84;1)$	0,7842	0,7842	0,7842
$EC_{nk}(0,84;1)$	0,9759	0,9759	0,9759
$PC_n(0,88;1)$	1,1295	1,4529	1,7764
$BC_n(0,88;1)$	0,8441	1,1245	1,4048
$CC_n(0,88;1)$	0,9099	1,2122	1,5145
$EC_n(0,88;1)$	1,4827	1,9949	2,5072
$PC_{nk}(0,88;1)$	0,8827	0,8827	0,8827
$BC_{nk}(0,88;1)$	0,7842	0,7842	0,7842
$CC_{nk}(0,88;1)$	0,7842	0,7842	0,7842
$EC_{nk}(0,88;1)$	0,9756	0,9756	0,9756

Tabla 4.13: Resultados de los ICP con una muestra simulada ($n=50$) mediante el modelo $PG2(\alpha_n, 1)$

α_n	$C_{pu}=1,00$	$C_{pu}=1,50$	$C_{pu}=2,00$
$PC_p(0,52;1)$	0,4808	0,6345	0,7881
$BC_p(0,52;1)$	0,8864	1,1697	1,4529
$CC_p(0,52;1)$	0,9616	1,2688	1,5761
$EC_p(0,52;1)$	0,9086	1,2203	1,5319
$PC_{pk}(0,52;1)$	0,3412	0,5206	0,5206
$BC_{pk}(0,52;1)$	0,8001	1,1179	1,1179
$CC_{pk}(0,52;1)$	0,9143	1,0304	1,0304
$EC_{pk}(0,52;1)$	0,8676	0,9989	0,9989
$PC_p(0,60;1)$	0,5463	0,7069	0,8676
$BC_p(0,60;1)$	0,8650	1,1195	1,3739
$CC_p(0,60;1)$	0,7305	0,7305	0,7305
$EC_p(0,60;1)$	1,0339	1,3888	1,7436
$PC_{pk}(0,60;1)$	0,3768	0,6012	0,6012
$BC_{pk}(0,60;1)$	0,7819	0,7819	0,7819
$CC_{pk}(0,60;1)$	0,7305	0,7305	0,7305
$EC_{pk}(0,60;1)$	0,9976	0,9976	0,9976
$PC_p(0,68;1)$	0,6215	0,8219	1,0223
$BC_p(0,68;1)$	0,8650	1,1195	1,3739
$CC_p(0,68;1)$	0,7305	0,7305	0,7305
$EC_p(0,68;1)$	1,1483	1,5473	1,9463
$PC_{pk}(0,68;1)$	0,4976	0,6822	0,6822
$BC_{pk}(0,68;1)$	0,7819	0,7819	0,7819
$CC_{pk}(0,68;1)$	0,7305	0,7305	0,7305

Tabla 4.13: Resultados de los ICP con una muestra simulada ($n=50$) mediante el modelo $PG2(\alpha_n, 1)$ (Continuación)

$EC_{nk}(0,68;1)$	0,9595	0,9595	0,9595
$PC_n(0,76;1)$	0,7013	0,9015	1,1017
$BC_n(0,76;1)$	0,8650	1,1195	1,3739
$CC_n(0,76;1)$	0,7305	0,7305	0,7305
$EC_n(0,76;1)$	1,2741	1,7180	2,1619
$PC_{nk}(0,76;1)$	0,5554	0,7605	0,7605
$BC_{nk}(0,76;1)$	0,7819	0,7819	0,7819
$CC_{nk}(0,76;1)$	0,7305	0,7305	0,7305
$EC_{nk}(0,76;1)$	0,9500	0,9500	0,9500
$PC_n(0,84;1)$	0,8017	1,0374	1,2729
$BC_n(0,84;1)$	0,8650	1,1195	1,3739
$CC_n(0,84;1)$	0,7305	0,7305	0,7305
$EC_n(0,84;1)$	0,8432	1,0862	1,3291
$PC_{nk}(0,84;1)$	0,6952	0,8428	0,8428
$BC_{nk}(0,84;1)$	0,7819	0,7819	0,7819
$CC_{nk}(0,84;1)$	0,7305	0,7305	0,7305
$EC_{nk}(0,84;1)$	0,7349	0,9796	0,9796
$PC_n(0,88;1)$	1,0813	1,4086	1,73598
$BC_n(0,88;1)$	0,8650	1,1195	1,3739
$CC_n(0,88;1)$	0,7305	0,7305	0,7305
$EC_n(0,88;1)$	0,8839	1,1383	1,3926
$PC_{nk}(0,88;1)$	0,8800	0,8800	0,8800
$BC_{nk}(0,88;1)$	0,7819	0,7819	0,7819
$CC_{nk}(0,88;1)$	0,7305	0,7305	0,7305
$EC_{nk}(0,88;1)$	0,8010	0,9824	0,9824

Tabla 4.14: Resultados de los ICP con una muestra simulada ($n=100$) mediante el modelo $PG2(\alpha_n, 1)$

α_n	$C_{pu}=1,00$	$C_{pu}=1,50$	$C_{pu}=2,00$
$PC_p(0,52;1)$	0,5048	0,6923	0,8798
$BC_p(0,52;1)$	0,9108	1,2489	1,5870
$CC_p(0,52;1)$	0,9602	1,3168	1,6734
$EC_p(0,52;1)$	0,9071	1,2187	1,5304
$PC_{pk}(0,52;1)$	0,4171	0,52	0,52
$BC_{pk}(0,52;1)$	0,8570	1,0545	1,0545
$CC_{pk}(0,52;1)$	0,9172	0,9172	0,9172
$EC_{pk}(0,52;1)$	0,8676	0,9939	0,9939
$PC_p(0,60;1)$	0,5368	0,7084	0,8801
$BC_p(0,60;1)$	0,7479	0,9871	1,2262
$CC_p(0,60;1)$	0,7879	1,0398	1,2917
$EC_p(0,60;1)$	1,0235	1,3783	1,7332
$PC_{pk}(0,60;1)$	0,4008	0,6014	0,6014
$BC_{pk}(0,60;1)$	0,71795	0,71795	0,71795
$CC_{pk}(0,60;1)$	0,6317	0,6317	0,6317
$EC_{pk}(0,60;1)$	0,9675	0,9675	0,9675
$PC_p(0,68;1)$	0,6181	0,8110	1,0040
$BC_p(0,68;1)$	0,7603	0,9976	1,2350
$CC_p(0,68;1)$	0,8009	1,0509	1,3009
$EC_p(0,68;1)$	1,1533	1,5523	1,9513
$PC_{pk}(0,68;1)$	0,48499	0,6821	0,6821
$BC_{pk}(0,68;1)$	0,7398	0,7876	0,7876
$CC_{pk}(0,68;1)$	0,6946	0,6946	0,6946

Tabla 4.14: Resultados de los ICP con una muestra simulada ($n=100$) mediante el modelo $PG2(\alpha_n, 1)$ (Continuación)

$EC_{nk}(0,68;1)$	0,9729	0,9729	0,9729
$PC_n(0,76;1)$	0,7034	0,9152	1,1271
$BC_n(0,76;1)$	0,7545	0,9818	1,2092
$CC_n(0,76;1)$	0,7948	1,0343	1,2737
$EC_n(0,76;1)$	1,2916	1,7355	2,1795
$PC_{nk}(0,76;1)$	0,5749	0,7602	0,7602
$BC_{nk}(0,76;1)$	0,7185	0,8025	0,8025
$CC_{nk}(0,76;1)$	0,7080	0,7080	0,7080
$EC_{nk}(0,76;1)$	0,9926	0,9926	0,9926
$PC_n(0,84;1)$	0,8019	1,0256	1,2494
$BC_n(0,84;1)$	0,8241	1,0540	1,2839
$CC_n(0,84;1)$	0,8884	1,1363	1,3842
$EC_n(0,84;1)$	0,8519	1,0948	1,3378
$PC_{nk}(0,84;1)$	0,6832	0,8413	0,8413
$BC_{nk}(0,84;1)$	0,7786	0,8819	0,8819
$CC_{nk}(0,84;1)$	0,8254	0,8254	0,8254
$EC_{nk}(0,84;1)$	0,7349	0,9992	0,9992
$PC_n(0,88;1)$	1,1850	1,4441	1,8033
$BC_n(0,88;1)$	0,8251	1,0403	1,2555
$CC_n(0,88;1)$	0,8895	1,1215	1,3536
$EC_n(0,88;1)$	0,8908	1,1451	1,3994
$PC_{nk}(0,88;1)$	0,8826	0,8826	0,8826
$BC_{nk}(0,88;1)$	0,7723	0,8923	0,8923
$CC_{nk}(0,88;1)$	0,8352	0,8352	0,8352
$EC_{nk}(0,88;1)$	0,8010	0,9973	0,9973

4.4.3. Resultados y Conclusiones

Entre los aspectos más resaltantes que pueden observarse en las tablas 4.12-4.14 se encuentran que, tanto el C_p como el C_{pk} estimados usando la función de densidad de la DPG, generan resultados muy por debajo de los valores fijados del C_{pu} , contrario a los índices estimados usando el método exacto, cuyos resultados son más próximos a dichos valores targets. En promedio se puede decir que para $C_{pu}=1$, (indica un proceso no capaz), los resultados promedio son: $PC_p= 0,711$; $BC_p= 0,850$; $CC_p=0,918$ y $EC_p=1,211$. Observándose que el método de percentil de Clements aproxima mejor al valor de C_{pu} en comparación con el método de percentil de Burr, mientras que el método exacto estima valores promedios por encima de 1. De manera similar ocurre con el estimado de C_{pk} . Esto es: $PC_{pk}= 0,5726$; $BC_p= 0,7876$; $CC_p=0,801$ y $EC_p= 0,9585$.

Cuando se fija $C_{pu}=1,5$ se tiene una situación similar, para los estimados de C_p ; $PC_p=0,931$; $BC_p= 1,134$, $CC_p=1,225$ y $EC_p=1,632$ y para el estimado de C_{pk} se obtuvo; $PC_{pk}=0,882$; $BC_p= 0,836$; $CC_p=0,8153$ y $EC_p=0,9768$, se observa en este caso que los valores están por debajo del valor target 1,5; pero sigue siendo el método exacto quien mejor estime los valores para los ICP.

Análogamente, para $C_{pu}=2,0$, se tiene que los estimados de C_p ; $PC_p=1,1511$; $BC_p= 1,4198$, $CC_p=1,523$ y $EC_p= 2,0484$ y para los estimados de C_{pk} se obtuvo; $PC_{pk}=0,7154$; $BC_p=0,836$; $CC_p=0,815$ y $EC_p=0,9768$. Esto indica que, posiblemente no es adecuado estimar los percentiles usando directamente la función de densidad PG, o en general cuando la asimetría en los datos es leve o moderada. Se observa además que a medida que el valor del parámetro, α_n se aproxima a 1, los estimadores de los ICP son más próximos a los valores targets. Otra situación que se pudo observar durante las corridas de los datos fue el hecho de que, para diferentes valores del parámetro α , los índices estimados no están definidos o simplemente generan índices negativos; por ejemplo, con α desde 0,53 hasta 0,59, los ICP resultan valores negativos. Similarmente entre 0,61 hasta 0,67 o 0,69 hasta 0,75; 0,77 hasta 0,83 y 0,85 hasta 0,87. De allí en adelante si es posible obtener resultados aceptables para los ICP. Es de hacer notar que la DPG cuando tiende a 1 su parámetro α , se aproxima a una exponencial, la cual representa una distribución con desviación severa de la normalidad. En general, el método exacto arroja resultados más próximos a los valores

targets lo que pudiera considerarse como mejor método de estimación. El programa que se construyó para lograr la información dada en las tablas 4.12 – 4.14 se encuentra en la sección de anexos (anexo 2).

www.bdigital.ula.ve

c.c Reconocimiento

CAPÍTULO 5

ÍNDICES DE CAPACIDAD DE PROCESOS BASADOS EN LA DPG3. APLICACIÓN

5.1. INTRODUCCION

Tal como se mencionó en el capítulo 3, distribuciones como Weibull, Beta, Gamma y Log-normal son utilizadas para investigar el efecto de datos no normales en la estimación de índices de capacidad de procesos, con la ventaja de que los valores de sus parámetros pueden estimarse usando diferentes métodos. Sin embargo cuando se habla de la distribución Pareto Generalizada se asocia a funciones de valor extremo con aplicaciones en datos hidrológicos, inundaciones, etc., mas no en aplicaciones de procesos productivos. Al estimar sus parámetros el investigador puede conseguirse con grandes limitantes por lo complejo del modelo, sobre todo cuando se refiere a la DPG con 3 y 4 parámetros. En este capítulo se construye un algoritmo genético siguiendo el procedimiento presentado en el capítulo 3, que permite estimar los parámetros de la DPG con 3 parámetros, usando el método de momento. Para ver la precisión de dicha estimación, se usa la idea de sesgo, varianza y cuadrado medio del error. Una vez que se logran estimar los parámetros, se simulan las muestras usando diferentes semillas aleatorias con la idea de estimar los ICP no normales usando los métodos descritos en las secciones anteriores. Por otro lado, se presentan algunas características de la DPG4 y se deja abierta la posibilidad de estudiarla en futuras investigaciones.

5.2 DISTRIBUCION GENERALIZADA DE PARETO CON TRES Y CUATRO PARÁMETROS

5.2.1 Distribución Pareto Generalizada con 3 parámetros

Una variante de la distribución Pareto Generalizada con 2 parámetros, la conforma la Distribución Pareto Generalizada con 3 parámetros, (DPG3).

Öztekin (2005), la presenta para una variable aleatoria Y con distribución exponencial estándar. Se define X como $X = \frac{\beta}{\alpha}(1 - \exp(-\alpha Y))$, donde α y β son los parámetros, para un

conjunto de n observaciones obtenidas de dicha variable. X sigue una distribución Pareto Generalizada con 3 parámetros si la función acumulada es de la forma.

$$F(x, \alpha, \beta, \gamma) = \begin{cases} 1 - \left(1 - \frac{\alpha}{\beta}(x - \gamma)\right)^{\frac{1}{\alpha}} & \alpha \neq 0 \\ \left(1 - \exp\left[\frac{x - \gamma}{\beta}\right]\right) & \alpha = 0 \end{cases} \quad (5.1)$$

para $x > \alpha$. Donde α es el parámetro de forma, $\beta > 0$ es el parámetro de escala, $\gamma > 0$ es el parámetro de localización.

Por otro lado, la función de densidad de probabilidad está dada por la siguiente expresión.

$$f(x, \alpha, \beta, \gamma) = \begin{cases} \frac{1}{\beta} \left[1 - \frac{\alpha(x - \gamma)}{\beta}\right]^{\left(\frac{1}{\alpha} - 1\right)} & \alpha \neq 0 \\ \frac{1}{\beta} \exp\left(-\frac{x - \gamma}{\beta}\right) & \alpha = 0 \end{cases} \quad (5.2)$$

para $\alpha < x < \infty$, $\beta > 0$ y $\gamma > 0$.

La expresión que define la función de los cuantiles de la DPG3 es:

$$x(F, \alpha, \beta, \gamma) = \gamma + \frac{\beta}{\alpha} \left[1 - (1 - F)^{\alpha}\right] \quad (5.3)$$

La DPG3 se reduce a una DPG con 2 parámetros cuando $\gamma=0$; es una distribución exponencial para $\alpha=0$ y $\gamma=0$ y es una distribución uniforme en $[0, \beta]$ para $\gamma=0$ y $\alpha=1$.

Los valores de la media, mediana y varianza están dados mediante las siguientes expresiones.

a. Media $\gamma + \frac{\beta}{1 + \alpha}$, con $\alpha < 1$,

b. Mediana, $\alpha + \frac{\beta(2^{\frac{1}{\alpha}} - 1)}{\gamma}$

c. Varianza, $\frac{\beta^2}{(1+\alpha)^2(1+2\alpha)}$, para $\alpha < \frac{1}{2}$

5.2.2 Distribución Pareto Generalizada con 4 parámetros

Elfattah et al. (2007), propusieron una distribución Generalizada de Pareto pero con 4 parámetros, (DPG4). El nuevo término representa un parámetro de forma, y la función de distribución de probabilidad está dada de la siguiente manera.

$$f(x, \alpha, \beta, \gamma, C) = \frac{C\alpha}{\beta} \left(\frac{x-\gamma}{\beta} \right)^{C-1} \left[\left(1 + \frac{x-\gamma}{\beta} \right)^C \right]^{-(\alpha+1)} \quad (5.4)$$

donde, $\gamma < x < \infty$, $\beta > 0$, $\alpha > 0$ y $C > 0$. En este caso, γ es el parámetro de localización, β es el parámetro de escala y (α, C) son parámetros de forma.

La función de distribución $F(x)$, está dada de la siguiente manera.

$$F(x, \alpha, \beta, \gamma, C) = 1 - \left[1 + \left(\frac{x-\gamma}{\beta} \right)^C \right]^{-\alpha} \quad (5.5)$$

La función para estimar los cuantiles está dada por:

$$x(F, \alpha, \beta, \gamma, C) = \gamma + \beta \left[1 - \frac{1}{(1-F)^{\frac{1}{\alpha}}} \right]^{\frac{1}{C}} \quad (5.6)$$

Además se tiene la función de supervivencia, $S(x)$ y la función de razón de Hazard, $H(x)$, la cual se utiliza mucho para cuantificar y caracterizar fenómenos de vida, sus expresiones respectivas se presentan a continuación;

$$S(x) = \left[\left(1 + \frac{x-\gamma}{\beta} \right)^C \right]^{-\alpha} \quad (5.7)$$

$$H(x) = \frac{C\alpha}{\beta} \left(\frac{x-\gamma}{\beta} \right)^{C-1} \left[\left(1 + \frac{x-\gamma}{\beta} \right)^C \right]^{-1} \quad (5.8)$$

Los parámetros de la DPG4 dados por la ecuación (5.4) pueden determinarse mediante lo siguiente.

a. Media,

$$\beta \frac{\Gamma\left(\alpha - \frac{1}{C}\right) \Gamma\left(1 + \frac{1}{C}\right)}{\Gamma(\alpha)} + \gamma$$

b. Varianza,

$$\beta^2 \left[\frac{\Gamma\left(1 + \frac{2}{C}\right) \Gamma\left(\alpha - \frac{2}{C}\right)}{\Gamma(\alpha)} - \left(\frac{\Gamma\left(1 + \frac{1}{C}\right) \Gamma\left(\alpha - \frac{1}{C}\right)}{\Gamma(\alpha)} \right)^2 \right]$$

c. Coeficiente de variación,

$$\left[\frac{\left[\Gamma\left(1 + \frac{2}{C}\right) \Gamma\left(\alpha - \frac{2}{C}\right) \Gamma(\alpha) - \Gamma^2\left(\alpha - \frac{1}{C}\right) \Gamma^2\left(\alpha - \frac{1}{C}\right) \right]^{\frac{1}{2}}}{\Gamma\left(1 + \frac{1}{C}\right) \Gamma\left(\alpha - \frac{1}{C}\right) + \gamma \Gamma(\alpha)} \right]$$

Elfattah, et al. (2007), exponen en un escrito que la distribución PG con 4-parámetros, puede transformarse en otras distribuciones conocidas tales como.

- i. En la ecuación (5.4), si $C=1$ y $\gamma=0$, la distribución PG se reduce a una DPG2.
- ii. Para $C=1$ en la ecuación (5.4) se obtiene una DPG3.
- iii. Cuando $C=1$ y $\gamma=0$, la ecuación (5.4) se reduce a una distribución Lomax con 2 parámetros y su función de densidad de probabilidad, está dada de la siguiente

$$\text{manera: } f(x, \alpha, \beta) = \frac{\alpha}{\beta} \left(1 + \frac{x}{\beta} \right)^{-(\alpha+1)} \quad \text{con } x > 0, \text{ y } \beta, \alpha > 0.$$

- iv. Cuando $C=1$, $\beta=1$ y $\gamma=0$, la ecuación (5.4) se reduce a una distribución Beta tipo II, cuya función de densidad de probabilidad es de la siguiente manera:
- $$f(x, \alpha) = \alpha(1+x)^{-(\alpha+1)} \quad \text{con } x > 0, \alpha > 0.$$

- v. Cuando $\beta=1$ y $\gamma=0$, la ecuación (5.4) se reduce a una distribución Burr XII, cuya función de densidad de probabilidad, está dada mediante la siguiente expresión:
- $$f(x, \alpha, C) = c\alpha x^{c-1} (1+x^c)^{-(\alpha+1)} \text{ con } x>0, \alpha>0 \text{ y } C>0.$$
- vi. Cuando $\beta^C=A$, la ecuación (5.4), se reduce a una distribución Gamma-Weibull compuesta por 4 parámetros y con función de densidad de probabilidad de la siguiente manera: $f(x, \alpha, A, \gamma, C) = C\alpha \left(\frac{x-\gamma}{A}\right)^{c-1} \left[\left(1 + \frac{x-\gamma}{A}\right)^c \right]^{-(\alpha+1)}$ con $\gamma < x < \infty$, $A > 0$, $\gamma > 0$ y $C > 0$.
- vii. Cuando $\beta^C=A$ y $C=2$, la ecuación (5.4) se reduce a una distribución Gamma-Raleigh con 3 parámetros, su función de densidad de probabilidad es como sigue:
- $$f(x, \alpha, A, \lambda) = 2\alpha \left(\frac{x-\gamma}{A}\right) \left[\left(1 + \frac{x-\gamma}{A}\right)^2 \right]^{-(\alpha+1)} \text{ con } \gamma < x < \infty \text{ y } A, \alpha > 0.$$
- viii. Cuando $\beta^C=A$ y $\gamma=1$, la ecuación (5.4) se reduce a una distribución exponencial-weibull con 3 parámetros. Su función de densidad de probabilidad está definida de la siguiente manera: $f(x, \alpha, A, \gamma, C) = C \left[\left(1 + \frac{x-\gamma}{A}\right)^c \right]^{-2}$ con $\alpha < x < \infty$ y $A, C > 0$.
- ix. Cuando $\beta^C=A$ y $C=1$, $\gamma=1$, la ecuación (5.4) se reduce a una distribución exponencial-exponencial con 2 parámetros y con función de densidad de probabilidad de la siguiente manera: $f(x, A, \gamma) = \frac{1}{A} \left(1 + \frac{x-\gamma}{A}\right)^{-2}$ con $\gamma < x < \infty$ y $A > 0$.

Además de estas propiedades anteriores, y haciendo cambio de variables más sofisticados, la ecuación (5.4), puede transformarse en otras distribuciones conocidas. Estas son:

- a. Cuando $y = \left[\ln \left[1 + \left(\frac{x - \gamma}{\beta} \right)^c \right] \right]^{\frac{1}{c}} + \gamma$, la ecuación (5.4), puede transformarse en una distribución Weibull, con función de densidad definida de la siguiente manera:
 $f(y; C, \alpha, \gamma) = C\alpha(y - \gamma)^{C-1} e^{-\alpha(y-\gamma)^C}$ con $\gamma < y < \infty$ y $C, \alpha > 0$.
- b. Cuando $y = \left[\ln \left[1 + \left(\frac{x - \gamma}{\beta} \right)^c \right] \right] + \gamma$, la ecuación (5.4) se puede transformar en una distribución exponencial con 2 parámetros. Su función de densidad de probabilidad está dada de la siguiente manera: $f(y; \alpha, \gamma) = \alpha e^{-\alpha(y-\gamma)}$, con $\gamma < y < \infty$ y $\alpha > 0$.
- c. Cuando se escribe, $y = \beta \left[\ln \left[1 + \left(\frac{x - \gamma}{\beta} \right)^c \right] \right] + \gamma$, la ecuación (5.4) puede transformarse en una distribución Logística generalizada tipo I con 3 parámetros y su función de densidad de probabilidad está dada como:
 $f(y; \alpha, \beta, \gamma) = \frac{\alpha}{\beta} e^{\left(\frac{y-\gamma}{\beta}\right)} \left[1 + e^{\left(\frac{y-\gamma}{\beta}\right)} \right]^{-(\alpha+1)}$ con $-\infty < y < \infty$.
- d. Cuando se escribe, $y = \left(\frac{x - \gamma}{\beta} \right)^c$ la ecuación (5.4) puede transformarse en una distribución Beta tipo II $\beta(1, \alpha)$, cuya función de densidad de probabilidad es como sigue: $f(y; \alpha) = \alpha [1 + y]^{-(\alpha+1)}$ con $y > 0$ y $\alpha > 0$.

5.3 ESTIMACION DE PARÁMETROS DE LA DGP3 MEDIANTE EL MÉTODO DE MÁXIMA VEROSIMILITUD

5.3.1 Método de Máxima Verosimilitud para la DPG con 3 parámetros.

Para el caso de la DPG3, se tomó en cuenta su función de densidad de probabilidad dada en la ecuación (5.2), para estimar sus parámetros: α (forma); β (escala) y γ (posición), usando el método de máxima verosimilitud.

Para una variable aleatoria X, la función de densidad de probabilidad de la DPG3 representada por la ecuación 5.2 es la siguiente:

$$f(x, \alpha, \beta, \gamma) = \begin{cases} \frac{1}{\beta} \left[1 - \frac{\alpha(x - \gamma)}{\beta} \right]^{\left(\frac{1}{\alpha} - 1\right)} & \alpha \neq 0 \\ \frac{1}{\beta} \exp\left(-\frac{x - \gamma}{\beta}\right) & \alpha = 0 \end{cases} \quad \text{para } \alpha < x < \infty, \beta > 0 \text{ y } \gamma > 0.$$

Con dicha ecuación se construye la función log-verosimilitud para la DPG3, que está dada por:

$$L(x_i, \alpha, \beta, \gamma) = -n \ln(\beta) + \left(\frac{1 - \alpha}{\alpha} \right) \sum_{i=1}^n \ln \left[1 - \frac{\alpha}{\beta} (x_i - \gamma) \right] \quad (5.9)$$

con x_i un valor muestral y n tamaño de la muestra.

Se calculan las expresiones de las derivadas parciales respecto a cada parámetro, lo que genera las siguientes ecuaciones.

$$\frac{\partial L}{\partial \alpha} = \frac{1}{\alpha^2} \sum_{i=1}^n \ln \left[1 - \frac{\alpha(x_i - \gamma)}{\beta} \right] + \frac{(1 - \alpha)}{\alpha} \sum_{i=1}^n \frac{(x_i - \gamma) / \beta}{1 - \frac{\alpha(x_i - \gamma)}{\beta}} \quad (5.10)$$

$$\frac{\partial L}{\partial \beta} = -\frac{n}{\beta} + \left(\frac{1 - \alpha}{\alpha} \right) \sum_{i=1}^n \left[\frac{\frac{\alpha(x_i - \gamma)}{\beta^2}}{1 - \frac{\alpha(x_i - \gamma)}{\beta}} \right] \quad (5.11)$$

$$\frac{\partial L}{\partial \gamma} = \left(\frac{1 - \alpha}{\alpha} \right) \sum_{i=1}^n \frac{\frac{\alpha}{\beta}}{1 - \frac{\alpha}{\beta} (x_i - \gamma)} \quad (5.12)$$

Las ecuaciones 5.10 hasta la 5.12 se igualan a cero, para obtener las soluciones necesarias. Para el caso del parámetro γ no existe un estimador de máxima verosimilitud, ya que la función de verosimilitud no está acotada respecto a γ . Sin embargo, Singh y Guo, (1995) y

Singh (1998), demostraron que se puede tomar el valor más pequeño de la muestra como estimado para γ .

Las ecuaciones (5.10) y (5.12), se hacen cero, cuando se cumple la siguiente igualdad

$$\sum_{i=1}^n \frac{\frac{(x_i - \gamma)}{\beta}}{1 - \frac{\alpha}{\beta}(x_i - \gamma)} = \frac{n}{(1 - \alpha)} \quad \text{y además} \quad \sum_{i=1}^n \ln \left[1 - \frac{\alpha}{\beta}(x_i - \gamma) \right] = -n\alpha$$

En general, los procedimientos para estimar los parámetros de la DPG3 usando este método de máxima verosimilitud se basan en la maximización de la ecuación (5.4). Los procedimientos pueden emplear el procedimiento de búsqueda de sección dorada (GSS siglas en inglés) descrita en Mathews, (1992); Press et al., (1992); Polak, (1997), Rheinboldt, (1998) otra manera es maximizar la ecuación (5.9) o usar el procedimiento de Newton-Raphson (Mathews, (1992); Press et al., (1992); Polak, (1997); Rheinboldt, (1998)). En cada caso se trata de resolver las ecuaciones, derivando parcialmente con respecto a cada parámetro desconocido, lo cual resulta bastante complejo, es por ello que en este trabajo se utilizó el método de algoritmo genético como alternativa para la estimación de los parámetros de la DPG3.

5.3.2 Método de momentos para la DPG con 3 parámetros

Siguiendo la metodología de Hosking y Wallis (1987), el método de momentos para

estimar los parámetros de la DPG3, se tiene que, $E\left(\frac{1 - \alpha(x - \gamma)}{\beta}\right)^r = \frac{1}{(1 + \alpha r)}$, si $(1 + \alpha r) > 0$.

Entonces el r-ésimo momento de X existe sí $\alpha > \frac{-1}{r}$. Los estimadores de momentos muestrales para la media (r=1), varianza (r=2) y asimetría (r=3), son los siguientes;

$$\bar{x} = \gamma + \frac{\beta}{1 + \alpha}, \quad S^2 = \frac{\beta^2}{(1 + \alpha)^2(1 + 2\alpha)} \quad \text{y} \quad \text{Asim} = \frac{2(1 - \alpha)(1 + 2\alpha)^{\frac{1}{2}}}{1 + 3\alpha}.$$

Para estimar los parámetros de la DGP3, usando el método de momentos, se resolvieron las ecuaciones generadas por la expresión, $E\left(\frac{1-\alpha(x-\gamma)}{\beta}\right)^r = \frac{1}{(1+\alpha r)}$, para el parámetro α y $r=1, 2,3,4$; luego se empleó el método de bisección (Mathews, 1992) y posteriormente se usan las ecuaciones de la media y la varianza para determinar β y γ , obteniéndose las expresiones para los estimadores de los parámetros siguientes:

$$\beta = S(1+\alpha)(1+2\alpha)^{\frac{1}{2}} \quad (5.13)$$

$$\gamma = \bar{x} - \frac{\beta}{1+\alpha} \quad (5.14)$$

5.4 DESEMPEÑO DEL METODO DE ESTIMACION DE MOMENTOS DE LA DPG3

De acuerdo a este método de momentos, se fijaron valores poblacionales para los parámetros α , β y γ presentados en la tabla 5.1; luego se estimaron los valores muestrales de dichos parámetros usando el software R, mediante las fórmulas 5.13 y 5.14; una vez obtenidos estos resultados (presentados en la tabla 5.2), se compararon con los poblacionales y escogieron los más próximos a ellos para así construir el modelo de la DPG3 y finalmente estimar los ICP con dicho modelo de distribución no normal.

Para evaluar el desempeño del método de momentos en la estimación de los parámetros de la DPG3, se tomó en cuenta el cuadrado medio del error; se condujo un experimento muestral de Monte Carlo, se simularon 10000 muestras aleatorias de tamaños 10, 20, 50, 75, 100 y 1000, se fijó un coeficiente de variación, ($Cv=0,5$) y se varió el coeficiente de asimetría, ($Asim=0,5$ y $2,5$).

Tabla 5.1: Casos poblacionales de la DPG3 considerados en el experimento muestral.

Distribución	Cv	Asim	α	β	γ
Caso 1	0,5	0,5	0,554	1,116	0,277
Caso 2	0,5	2,5	-0,069	0,433	0,536

Elaboración: Fuente propia

Caso 1: Cuando el coeficiente de variación y asimetría son respectivamente 0,5 y 0,5; se fijan los valores poblacionales de $\alpha=0,554$; $\beta=1,116$ y $\gamma=0,277$.

Tabla 5.2: Valores estimados de la distribución GP3 para cada conjunto de parámetros $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ ($Cv=0,5$ y $Asim=0,5$)

n	10	20	50	75	100	1000
$\hat{\alpha}$	0,9258	0,8766	0,6939	0,6715	0,6473	0,5706
$\hat{\beta}$	1,4306	1,4283	1,2623	1,2428	1,2136	1,1298
$\hat{\gamma}$	0,2902	0,3014	0,2890	0,2425	0,2861	0,2804

Elaboración: Fuente propia

Después de simular los datos y hallar los parámetros estimados, se puede observar que a medida que el tamaño de la muestra aumenta, los valores muestrales se aproximan más a los valores poblacionales por lo que se asume que estos se distribuyen normalmente. Este supuesto permitió tomar las 10000 repeticiones de cada estimación paramétrica y calcular el cuadrado medio del error, (CME), o el sesgo elevado al cuadrado más la varianza de la distribución de los estimados paramétricos. El sesgo se define como la diferencia entre el valor del parámetro verdadero y el estimado del parámetro. Matemáticamente es, $E(\hat{\theta}) - \theta$, donde $\hat{\theta}$ es el parámetro estimado y θ , es el valor verdadero del parámetro. En este caso, el sesgo se calcula hallando el promedio de cada uno de los 10000 parámetros estimados, restado del verdadero valor usado para generar los datos, esto es:

$$sesgo = \frac{1}{n} \sum_{i=1}^n x_i - certeza$$

En la tabla 5.3, se presentan los sesgos calculados para cada conjunto de parámetros y tamaños muestrales. Así como en la figura 5.1, se presentan las gráficas obtenidas para la DPG3 con el conjunto de parámetros $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ (Coeficiente de variación=0,5 y asimetría = 0,5) para la muestra simulada tamaño 10000.

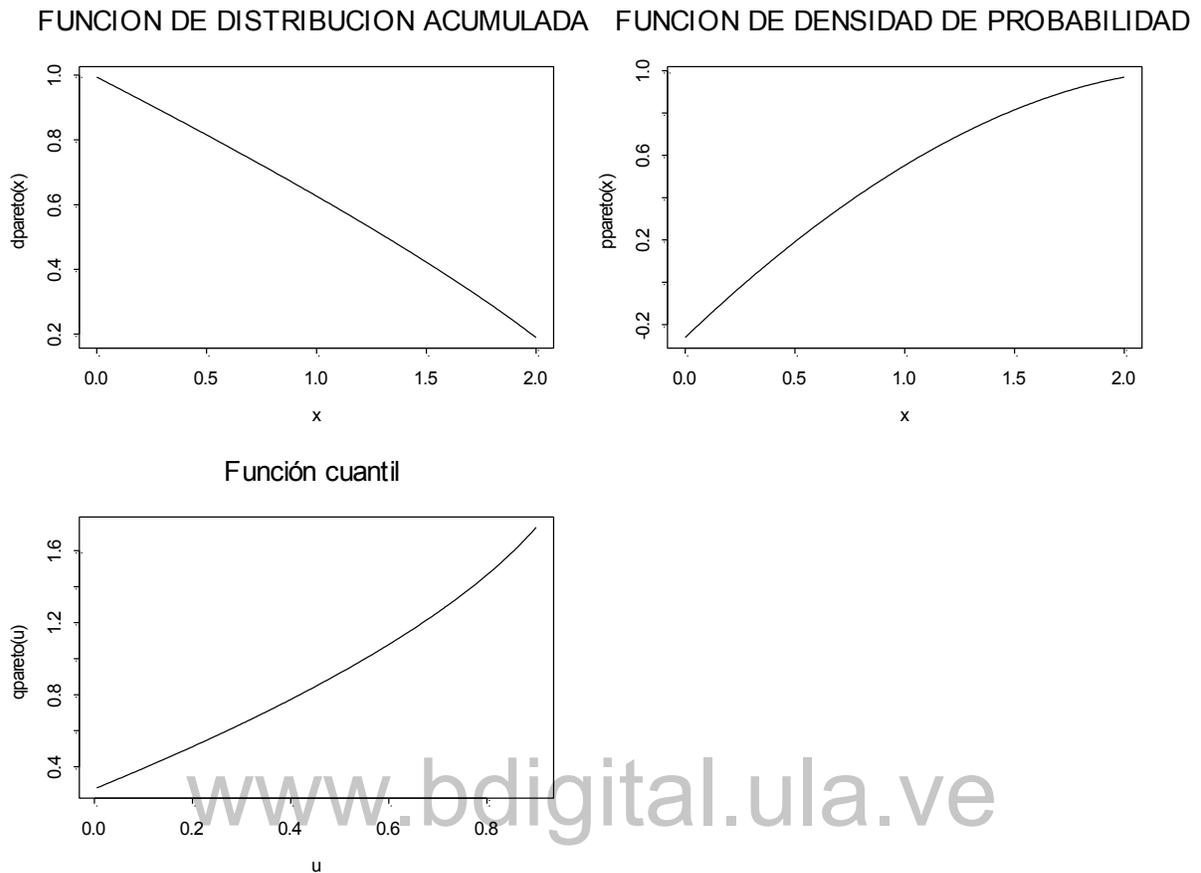


Figura 5.1: Función DPG3 el conjunto de parámetros $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$

Tabla 5.3: Sesgo para cada conjunto de parámetros $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ ($Cv=0,5$ y $Asim=0,5$)

n	10	20	50	75	100	1000	5000	10000
$\hat{\alpha}$	0,3718	0,3226	0,1399	0,1175	0,0933	0,0166	0,00865	0,0052
$\hat{\beta}$	0,3146	0,3123	0,1463	0,1268	0,0976	0,0138	0,01085	0,0055
$\hat{\gamma}$	0,0132	0,0244	0,0120	0,0155	0,0091	0,0034	0,000163	0,00069

Elaboración: Fuente propia

El sesgo del estimador de momentos se aproxima a cero a medida que el tamaño muestral n se hace más grande.

Por otro lado, la varianza, está definida para ser la desviación de la media. Se calculó usando la suma de las desviaciones del cuadrado de cada punto del valor promedio de los

datos y dividiéndola por $n-1$, esto es, $varianza = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \left(\frac{1}{n-1} \sum_{i=1}^n x_i \right)^2 \right]$. Los resultados de dichas varianzas se presentan en la tabla 5.4.

Tabla 5.4: Varianza para el conjunto de parámetros $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ ($Cv=0,5$ y $Asim=0,5$)

n	10	20	50	75	100	1000	5000	10000
Varianza	0,1951	0,2073	0,2313	0,2360	0,2359	0,2414	0,2446	0,2441

Elaboración: Fuente propia

Una vez que se tienen las medidas de sesgo y varianza se pueden calcular el cuadrado medio del error usando la expresión siguiente: $CME = sesgo^2 + varianza$

En la tabla 5.5, se muestran los cuadrados medios para el conjunto de parámetros $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ ($Cv=0,5$ y $Asim=0,5$).

Tabla 5.5: Cuadrado medio del error para el conjunto de parámetros $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$ ($Cv=0,5$ y $Asim=0,5$)

n=	10	20	50	75	100	1000	5000	10000
$CME_{\hat{\alpha}}$	0,3334	0,3114	0,2509	0,2498	0,2446	0,2417	0,2447	0,2442
$CME_{\hat{\beta}}$	0,2941	0,3049	0,2528	0,2521	0,2455	0,2416	0,2447	0,2442
$CME_{\hat{\gamma}}$	0,1953	0,2079	0,2315	0,2362	0,2360	0,2414	0,2446	0,2441

Elaboración: Fuente propia

Caso 2: Cuando el coeficiente de variación y asimetría son respectivamente 0,5 y 2,5; se fijan los valores poblacionales de $\alpha=-0,069$; $\beta=0,433$ y $\gamma=0,536$; se tienen los resultados presentados en la tabla 5.6.

Tabla 5.6 Valores estimados de la DGP3 para cada conjunto de parámetros $\alpha=-0,069$; $\beta=0,433$ y $\gamma=0,536$ ($Cv=0,5$ y $Asim=2,5$)

n	10	20	50	75	100	1000	5000	10000
$\hat{\alpha}$	0,4275797	0,31836	0,1337	0,066 6	0,0700	-0,0192	-0,0463	-0,050
$\hat{\beta}$	0,6718496	0,60480	0,5391	0,5070	0,5043	0,461	0,4454	0,4441
$\hat{\gamma}$	0,4701308	0,47877	0,5049	0,5271	0,5144	0,5258	0,5309	0,5313

Elaboración: Fuente propia

En la tabla 5.7, se presentan los sesgos para cada conjunto de parámetros estimados y tamaños muestrales.

Tabla 5.7: Sesgo para cada conjunto de parámetros $\alpha=-0,069$; $\beta=0,433$ y $\gamma=0,536$ ($Cv=0,5$ y $Asim=0,5$)

n	10	20	50	75	100	1000	5000	10000
$\hat{\alpha}$	0,4965	0,3873	0,2027	0,1355	0,1390	0,0497	0,2405	0,0189
$\hat{\beta}$	0,2388	0,1718	0,1061	0,0741	0,0713	0,0280	0,01243	0,0111
$\hat{\gamma}$	-0,065	-0,0572	-0,0310	-0,0088	-0,0215	-0,0101	-0,0050	-0,0046

Elaboración: Fuente propia

Tanto en el caso 1 ($Cv=0,5$ y $asimetría=0,5$), como en el caso 2 ($Cv=0,5$ y $asimetría=2,5$) el sesgo del método de momentos se aproxima a cero a medida que el tamaño muestral n se hace más grande. Lo que demuestra la tendencia a la asintoticidad insesgada. Cabe señalar que en ambos casos, la varianza global se acerca al valor de 0,24.

5.5. APLICACIÓN DEL ALGORITMO GENÉTICO PARA ESTIMAR LOS PARÁMETROS DE LA DPG3

En esta sección, se presenta el procedimiento seguido para estimar los valores de los parámetros muestrales considerando las 10000 muestras simuladas anteriormente. Los valores poblacionales considerados en este caso son, $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$. Una vez obtenido los estimados de los 3 parámetros, el modelo de la DPG3 queda definido y finalmente queda estimar los ICP no normales.

Se usó R y un algoritmo genético cuyos programas se encuentran en la sección de anexos (anexo 5).

5.5.1. Sistema de ecuaciones no lineales, escrito en el software R.

Las siguientes ecuaciones definen la DPG3, según el método de momentos: $\bar{x} = \gamma + \frac{\beta}{1+\alpha}$,

$$S^2 = \frac{\beta^2}{(1+\alpha)^2(1+2\alpha)} \quad \text{y} \quad Asim = \frac{2(1-\alpha)(1+2\alpha)^{\frac{1}{2}}}{1+3\alpha}, \text{ las cuales permitieron estimar los}$$

valores de los parámetros usando R. Usando el software R, estas ecuaciones anteriores se escribieron de la siguiente manera:

$x[1]$ que denota el parámetro α

$x[2]$ representa el parámetro β

$x[3]$ representa el parámetro γ .

$$x[3] + (x[2])/(1 + x[1]) = \text{Media}[i]$$

$$(x[2]^2)/(1 + 2*x[1])*(1 + x[1])^2 = \text{varianza}[i]$$

$$2*(1 - x[1])*sqrt(1 + 2*x[1])/(1 + 3*x[1]) = \text{Skewness}[i]$$

Se construyó un sistema de ecuaciones no lineales usando el método de bisección y considerando $\bar{x}=0,9771234$; $\text{varianza}= 0,2338198$ y $\text{Skewness}[i]= 0,5645426$ y finalmente se obtuvieron los resultados para los estimados de los parámetros: $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$.

Valores obtenidos usando el software R para la primera muestra

Media[1]	0,9771234
Varianza[1]	0,2338198
Skewness[1]	0,5645426

Elaboración: Fuente propia

5.5.2. Algoritmo genético.

Siguiendo los pasos planteados en el capítulo 3, se describe a continuación los pasos seguidos para la construcción del algoritmo genético en la búsqueda de los parámetros del modelo de la DPG3.

- **Inicialización:** La población inicial está constituida por los tres parámetros que definen la DPG3 (Fórmula 5.1), los cuales son: $\alpha=0,554$, $\beta=1,116$ y $\gamma=0,277$, esto define un cromosoma como un vector fila de la forma $[0,554; 1,116; 0,277]$ que será representado en un sistema binario.

Así el modelo estará constituido de la siguiente manera,

$$F(x, 0,554, 1,116, 0,277) = \begin{cases} 1 - \left(1 - \frac{0,554}{1,116}(x - 0,277)\right)^{\frac{1}{0,554}} & \alpha \neq 0 \\ \left(1 - \exp\left[\frac{x - 0,277}{1,116}\right]\right) & \alpha = 0 \end{cases}$$

con, $\alpha \neq 0$, $\beta > 0$, $\gamma > 0$.

- Función de aptitud: Se utilizaron las fórmulas para la media, la varianza y la asimetría generadas con el método de momentos, $\bar{x} = \gamma + \frac{\beta}{1 + \alpha}$, $S^2 = \frac{\beta^2}{(1 + \alpha)^2 (1 + 2\alpha)}$ y

$Asim = \frac{2(1 - \alpha)(1 + 2\alpha)^{\frac{1}{2}}}{1 + 3\alpha}$, como funciones de aptitud para generar los valores de los momentos muestrales.

- Condición de término: Se realizaron diferentes corridas de diferentes tamaños, pero a partir de las 500 iteraciones no se observaron cambios en la población, se fijó la precisión en 0,000001. Las iteraciones fueron disminuyendo luego de 10 en 10, comprobando que cada valor generado con las formulas convergieran a un valor común. Por otro lado el número de individuos seleccionados fueron 8, los cuales fueron suficientes para obtener los valores estimados de α_i , β_i , γ_i , con $i=1,2,\dots,8$. Se consideró una probabilidad de ocurrencia de 0,8; esto garantizó la selección de los cromosomas más próximos a las soluciones iniciales que resultaron ser 4. Se mutaron estos 4 individuos entre sí y se transformaron a sus respectivas representaciones binarias.

En la tabla 5.8, se recoge un conjunto de valores donde se observan los resultados promedios obtenidos para los parámetros estimados usando el software R y algoritmo genético. El resultado obtenido con el algoritmo genético fue promedio obtenido para cada 8 valores de cada momento muestral ya que este no arrojó un valor óptimo para cada parámetro, pero si fue mejor valor en comparación con el obtenido con el otro método.

Tabla 5.8: Resultados obtenidos usando Algoritmo genético y el software R.

Parámetro	Usando	Algoritmo	Número de	tiempo(m)	precisión de
$\hat{\alpha}$	0,5007487	0,5007490	500	1,85052	4,9 E-7
$\hat{\beta}$	1,0266590	1,0266590	508	0,28463	1,8E-7
$\hat{\gamma}$	0,2930257	0,2949120	500	1,39218	5,2 E-7

Elaboración: Fuente propia

5.6 METODOS PARA ESTIMAR LOS INDICES DE CAPACIDAD DE PROCESOS USANDO LA DPG3

En esta sección del trabajo, se introduce a la DPG3 como modelo para la estimación de ICP. Se utilizó su función de densidad dada por la fórmula 5.1, la cual se sustituyó en las expresiones correspondientes a los índices C_p y C_{pk} no normales presentadas en el capítulo 2, (ecuaciones 2.2 – 2.5), generando así una nueva manera de estimar los ICP antes mencionados, usando exclusivamente la función de densidad acumulada de la DPG3. A continuación se define un método utilizado por Castagliola (1996), que ilustra como estimar los ICP usando la función de densidad de cualquier distribución no normal.

5.6.1. Método ICP usando una función de distribución acumulada

Wierda (1993), introdujo un enfoque para evaluar capacidad de procesos para datos no normales usando una función de distribución acumulada. Castagliola (1996), usó este enfoque para calcular la proporción de ítems no conformes y estimar los índices de capacidad usando dicha proporción. Mostró, además la relación entre capacidad del proceso y la proporción de ítems no conformes y usó el método de la función acumulada para evaluar los ICP considerando datos no normales ajustando una distribución de Burr a los datos del proceso.

Usando el método de Castaglioga (1996), los índices C_p y C_{pk} se definen como:

$$C_p = \frac{\Phi^{-1}(0.5+0.5 \int_{lei}^{les} f(x)dx)}{3} \quad (5.15)$$

$$C_{pk} = \min(C_{ps}, C_{pi}) \quad (5.16)$$

donde,

$$C_{pi} = \frac{\phi^{-1}\left(0.5+0.5 \int_{lei}^T f(x)dx\right)}{3} \quad (5.17)$$

$$C_{ps} = \frac{\phi^{-1}\left(0.5+0.5 \int_T^{les} f(x)dx\right)}{3} \quad (5.18)$$

con, $f(x)$ como la función de densidad de probabilidad del proceso y T representa la media del proceso para datos normales y mediana del proceso para datos no normales.

5.6.2. Estimación de los ICP usando la Función de densidad de probabilidad

En esta sección, se explica cómo se pueden estimar los ICP usando la función de densidad de la DPG3. Siguiendo la metodología usada por Liu y Chen (2006) para estimar los ICP no normales es necesario estimar los valores de los percentiles del modelo Pareto Generalizado.

Tal como se expresó en el capítulo 2, sección 2.3.4, para estimar los valores de estos percentiles denotados $x_p = F_X^{-1}(p)$ se utiliza la expresión $\hat{x}_p = \hat{F}_X^{-1}(p)$, tomando en cuenta que la función puede o no ser continua lo que implica que puede o no existir solución. Cuando F_X es estrictamente creciente y continua, $F_X^{-1}(p)$ es un valor único, pero cuando F_X no es estrictamente creciente o continua, puede existir más de una solución o posiblemente ninguna.

De las fórmulas 2.2 y 2.19, se obtiene la siguiente expresión:

$$C_p = \frac{U_t - L_t}{x(F_{99,865}) - x(F_{0,135})} = \frac{[U_t - L_t]}{\frac{\beta}{\alpha} \left[(1 - F_{0,135})^\alpha - (1 - F_{99,865})^\alpha \right]} \quad (5.19)$$

Por otro lado, de las fórmulas 2.3 hasta 2.4 se tiene lo siguiente:

$$C_{ps} = \frac{U_t - \text{Med}}{U_p - \text{Med}} = \frac{U_t - \left[\gamma + \frac{\beta}{\alpha} \left[1 - (1 - F_{50})^\alpha \right] \right]}{\frac{\beta}{\alpha} \left[(1 - F_{50})^\alpha - (1 - F_{99,865})^\alpha \right]} \quad (5.20)$$

$$C_{pi} = \frac{\text{Med} - L_t}{\text{Med} - L_p} = \frac{\gamma + \frac{\beta}{\alpha} \left[1 - (1 - F_{50})^\alpha \right] - L_t}{\frac{\beta}{\alpha} \left[(1 - F_{50})^\alpha - (1 - F_{99,865})^\alpha \right]} \quad (5.21)$$

Finalmente,

$$C_{pk} = \text{Min}(C_{pu}, C_{pi}) \quad (5.22)$$

5.6.3 Índices de capacidad basados en el método exacto considerando la DPG3

Siguiendo un procedimiento similar a lo antes planteado, se consideró la función de distribución de probabilidad de la DPG3, se sustituyó en las fórmulas 2.14-2.16, usadas en el método exacto, lográndose lo siguiente:

$$\gamma + \frac{\beta}{\alpha} \left[1 - (1 - 0,00135)^\alpha \right] = X_{0,00135} \quad (5.23)$$

$$\gamma + \frac{\beta}{\alpha} \left[1 - (1 - 0,50)^\alpha \right] = X_{0,50} \quad (5.24)$$

$$\gamma + \frac{\beta}{\alpha} \left[1 - (1 - 0,99865)^\alpha \right] = X_{0,99865} \quad (5.25)$$

5.7. ESTIMACION DE ICP USANDO UNA VARIABLE DPG3. APLICACIÓN

Tomando en cuenta el proceso de estimación de los parámetros de una DPG3 mediante el método de momentos, se pueden generar muestras que sigan dicha distribución.

Para estudiar el desempeño de los índices de Capacidad de Procesos, se llevó a cabo un conjunto de simulaciones tomando en cuenta una muestra aleatoria de

tamaño $n=100$ elementos, con distribución es Pareto Generalizada con 3 parámetros, los cuales son: $\hat{\alpha}= 0,554$; $\hat{\beta} = 1,116$ y $\hat{\gamma} = 0,277$. Se usó una semilla y se obtuvo los resultados que se resumen en las tablas 5.9 y 5.10.

Tabla 5.9: Resumen estadístico para la DPG con $n=100$ y el conjunto de parámetros:
 $\hat{\alpha}= 0,554$; $\hat{\beta} = 1,116$ y $\hat{\gamma} = 0,277$.

Min	1er cuartil	Mediana	3er cuartil	Max	Media	Desviación	Asimetría	Kurtosis
0,279	0,53	0,86	1,31	2,19	0,943	0,490	0,58	-0,526

Elaboración: Fuente propia

Tabla 5.10: Valores de los Cuantiles de la DPG(0,554; 1,116; 0,277)

$q_{0,00135}$	$q_{0,5}$	$q_{0,95}$	$q_{0,99865}$	L_p	U_p	L_t
0,279	0,919	1,91	2,24	0,283	2,18	0,43

Elaboración: Fuente propia

5.7.1. Resultados de la estimación de ICP usando métodos no normales.

De acuerdo a los resultados presentados en las tablas 5.9 y 5.10 se resumen en las tablas 5.11; 5.12; 5.13 y 5.14 los valores de los ICP estimados usando el modelo de la DPG3(0,554; 1,116; 0,277).

Tabla 5.11: Valores de los ICP usando la DPG(0,554; 1,116; 0,277), método de Clements para diferentes valores de U_t .

	$C_{pu}=0,5$	$C_{pu}=1,0$	$C_{pu}=2,0$
	$U_t=1,52$	$U_t=2,18$	$U_t=3,49$
C_p	0,57	0,921	1,61
C_{pks}	0,499	1,0	1,994
C_{pki}	0,744	0,744	0,744

Elaboración: Fuente propia

La metodología de estimar los ICP propuesta por Liu-Chen (2006) en su artículo, depende de los valores de la asimetría y kurtosis de la muestra para conseguir $\alpha_3=0,57$ y $\alpha_4=2,4$. Con estos se determinan los puntos percentiles estandarizados que se encuentran tabulados en dicho artículo.

Para este caso se tienen los siguientes puntos percentiles.

Percentiles estandarizados de Burr: $BZ_{0,00135} = -1,257$; $BZ_{0,5} = -0,141$ y $BZ_{0,99865} = 3,121$.

Percentiles estandarizados de Clements: $CZ_{0,00135} = -1,539$; $CZ_{0,5} = -0,142$ y $CZ_{0,99865} = 2,702$

A continuación se presenta las tablas 5.12 y 5.13 con los valores de C_p , C_{pks} y C_{pki} estimados con el método de percentiles de Burr y percentiles de Clements respectivamente.

Tabla 5.12: Valores de los ICP usando la DPG(0,554; 1,116; 0,277), método de Percentil de Burr para diferentes valores de U_t .

	$C_{pu}=0,5$	$C_{pu}=1,0$	$C_{pu}=2,0$
	$U_t=1,52$	$U_t=2,18$	$U_t=3,49$
C_p	0,5081	0,816	1,426
C_{pks}	0,406	0,610	1,221
C_{pki}	0,806	0,806	0,806

Elaboración: Fuente propia

Tabla 5.13: Valores de los ICP usando la DPG(0,554; 1,116; 0,277), método de Percentil de Clements para diferentes valores de U_t .

	$C_{pu}=0,5$	$C_{pu}=1,0$	$C_{pu}=2,0$
	$U_t=1,52$	$U_t=2,18$	$U_t=3,49$
C_p	0,525	0,842	1,473
C_{pks}	0,524	0,997	1,937
C_{pki}	0,527	0,527	0,527

Elaboración: Fuente propia

Tabla 5.14: Valores de los ICP usando la DPG(0,554; 1,116; 0,277), método Exacto para diferentes valores de U_t .

	$C_{pu}=0,5$	$C_{pu}=1,0$	$C_{pu}=2,0$
	$U_t=1,52$	$U_t=2,18$	$U_t=3,49$
C_p	0,556	0,891	1,36
C_{pks}	0,455	0,953	1,95
C_{pki}	0,764	0,764	0,764

Elaboración: Fuente propia

En los resultados presentados en las tablas 5.9 – 5.14, puede observarse que el método de Clements y el de percentil de Clements generan resultados para el C_p como el C_{pks} más próximos a los valores target de C_{pu} considerando muestras de diferentes tamaños; es de hacer notar que sólo cuando $U_t=2$, se puede lograr un índice que define un proceso estadísticamente capaz; esto permite decidir, con cuál método se puede estimar mejor los ICP usando la DPG3.

Otro resultado que se estimó, dada la muestra simulada, fueron los ICP suponiendo que la distribución de los datos es Normal; estos son presentados en la tabla 5.15.

Tabla 5.15: Valores de los ICP usando el modelo $N(0,94; 0,49)$ para diferentes valores de U_t .

	$C_{pu}=0,5$	$C_{pu}=1,0$	$C_{pu}=2,0$
	$U_t=1,52$	$U_t=2,18$	$U_t=3,49$
C_p	0,371	0,595	$C_p=1,04$
C_{pks}	0,349	0,84	$C_{pks}=1,74$
C_{pki}	0,349	0,349	$C_{pki}=0,349$

Elaboración: Fuente propia

Es de hacer notar la real diferencia tanto en el C_p como en el C_{pks} al considerar una DPG(0,554; 1,116; 0,277), y una distribución normal, la cual genera ICP no capaz, aun cuando lo sean usando el modelo de la DPG3, lo que puede conllevar a quienes utiliza estos ICP a tomar decisiones erróneas respecto a un proceso productivo.

A continuación se presentan algunas características de manera resumida la DPG con 4 parámetros, la cual puede utilizarse para estimar los ICP usando el método exacto y la función cuantil generada por dicho modelo. Es de hacer notar que para la estimación de los parámetros de esta distribución, se amerita el conocimiento de análisis numéricos para resolver ecuaciones no lineales generadas, las cuales resultan muy engorrosas de trabajar. Se propone para futuras investigaciones. Sin embargo, se indica un conjunto de expresiones que pueden ayudar a estimar los ICP usando el método exacto y la función de densidad de probabilidad de la DPG4.

5.8. ÍNDICES DE CAPACIDAD DE PROCESOS BASADOS EN EL MÉTODO EXACTO TOMANDO EN CUENTA LA DPG4.

En este caso se tienen las siguientes expresiones.

$$\gamma + \beta \left[1 - \frac{1}{(1 - F_{0,135})^{\frac{1}{\alpha}}} \right]^{\frac{1}{C}} = X_{0,00135} \quad (5.26)$$

$$\gamma + \beta \left[1 - \frac{1}{(1 - F_{0,50})^{\frac{1}{\alpha}}} \right]^{\frac{1}{C}} = X_{0,50} \quad (5.27)$$

$$\gamma + \beta \left[1 - \frac{1}{(1 - F_{0,99865})^{\frac{1}{\alpha}}} \right]^{\frac{1}{C}} = X_{0,99865} \quad (5.28)$$

5.9 ÍNDICES DE CAPACIDAD BASADOS EN LA FUNCIÓN DE PROBABILIDAD DE LA DPG4

A continuación se presentan algunas expresiones de los ICP estimados mediante el modelo de la DPG4; estos se dedujeron de la fórmula (2.1) y la fórmula de la función de

densidad de probabilidad $x(F, \alpha, \beta, \gamma, C) = \gamma + \beta \left[1 - \frac{1}{(1 - F)^{\frac{1}{\alpha}}} \right]^{\frac{1}{C}}$

Son presentados como alternativa abierta para futuros trabajos de investigación. Se obtienen de la siguiente manera:

$$C_p = \frac{U - L}{x(F_{99,865}) - x(F_{0,135})} = \frac{U - L}{\beta \left[\left[1 - \frac{1}{(1 - F_{99,865})^{\frac{1}{\alpha}}} \right]^{\frac{1}{C}} - \left[1 - \frac{1}{(1 - F_{0,135})^{\frac{1}{\alpha}}} \right]^{\frac{1}{C}} \right]} \quad (5.29)$$

Para estimar el C_{pk} usando la DG4 se consideran las fórmulas (2.2)-(2.4) y la ecuación de la función cuantil, lográndose lo siguiente.

$$C_{pu} = \frac{U - M}{U_p - M} = \frac{U - \left[\gamma + \beta \left[1 - \frac{1}{(1 - F_{50})^{\frac{1}{\alpha}}} \right]^{\frac{1}{c}} \right]}{\beta \left[\left[1 - \frac{1}{(1 - F_{99,865})^{\frac{1}{\alpha}}} \right]^{\frac{1}{c}} - \left[1 - \frac{1}{(1 - F_{50})^{\frac{1}{\alpha}}} \right]^{\frac{1}{c}} \right]} \quad (5.30)$$

$$C_{pl} = \frac{M - L_t}{M - L_p} = \frac{\gamma + \beta \left[1 - \left[1 - \frac{1}{(1 - F_{50})^{\frac{1}{\alpha}}} \right]^{\frac{1}{c}} \right] - L}{\beta \left[\left[1 - \frac{1}{(1 - F_{50})^{\frac{1}{\alpha}}} \right]^{\frac{1}{c}} - \left[1 - \frac{1}{(1 - F_{0,135})^{\frac{1}{\alpha}}} \right]^{\frac{1}{c}} \right]} \quad (5.31)$$

Luego, $C_{pk} = \text{Min}(C_{pu}, C_{pl})$

CONCLUSIONES

El método de percentiles de Burr propuesto por Liu y Chen (2006), arroja resultados ligeramente mejores comparados con el método de percentiles de Clements, esto es, valores más próximos a los valores targets fijados para el C_{pu} {0,50; 1,0; 2,0}, con menor dispersión en las distribuciones con asimetría severa y más aún para una DPG2; esto permite suponer que, en cuanto a precisión, los ICP estimados para las distribuciones con asimetrías moderadas arrojan resultados menos confiables.

Por otro lado, de acuerdo a los datos facilitados por la empresa productora de alimento ubicada en Cumaná Estado Sucre, ajustados con la DPG2, se observó que los índices obtenidos con el método de percentiles de Clements y los obtenidos con el método de Burr producen mejores resultados, más cercanos al valor target, mientras que, los índices calculados suponiendo normalidad en los datos resultaron mucho más pequeños y por lo tanto más alejado de los valores target de C_{pu} . Esto indica que posiblemente sí influye la asimetría de la distribución respecto a su normalidad sobre todo cuando esta es severa, por lo que se recomienda el método de Percentiles de Burr para estimar los ICP.

Otro hecho que se pudo determinar en este estudio, fue la diferencia lograda al comparar los valores de los ICP obtenidos usando la función de probabilidad de la DPG2 para determinar directamente los valores de los percentiles 99,865; 0,50 y 0,035. Los resultados logrados fueron casi siempre menores que los ICP usando los métodos de percentiles de Clements, de Burr y método exacto. Sin embargo, el método exacto proporcionó los mejores resultados. Por lo que puede decirse que el método exacto tiende a ser mejor.

Tomando en cuenta un modelo no normal y el porcentaje de cubrimiento generado por los intervalos bootstrap, se puede generalizar que de los tres tipos de método utilizados para evaluar los índices de capacidad de procesos no normales, el método de percentil de Burr arrojó mejores resultados que el método de percentil de Clements. Mientras que, las dispersiones entre los índices no normales se mantienen significativamente iguales; por el contrario existe una variabilidad notable entre los índices estimados no normales y normales. Otro detalle que resalta en las estimaciones del índice C_{pk} normal es que éste mantiene constante para los diferentes valores targets fijados tanto para el C_{pu} como U; esto

posiblemente porque la media de la muestra es mayor que la mediana y al evaluarlos siempre se va a tomar el valor más pequeño.

Cuando la asimetría de la distribución de los datos es severa, tiende a generarse resultados erróneos al evaluar el desempeño de los ICP, esto se pudo constatar sobre todo cuando fueron estimados tomando en cuenta los métodos de Percentiles de Clements y de Burr. El índice estimado usando el método de Clements aplicado a las distribuciones Weibull, Beta, Log-normal y PG, arrojó el siguiente orden de estimación, PG, luego Weibull, siguiendo la distribución Beta y finalmente Log-normal, lo que indica que mientras más severa es la desviación en la distribución normal, el índice estimado con el método de Clements generan resultados más próximos a los valores targets, en comparación con los estimados usando el método de Burr.

En cuanto a la precisión, las distribuciones con asimetrías moderadas son menos precisas; un orden de las distribuciones, de menos asimétrica a mas, es el siguiente; Log-normal, siguiendo la Beta, luego Weibull y finalmente PG. Los resultados del índice C_{pk} estimado mediante el método de Percentil de Burr en comparación con el método de Clements, siempre fueron mayores. Finalmente, de todas las distribuciones no normales consideradas, la DPG2(0,1; 0,2) arrojó las desviaciones más pequeñas usando ambos métodos, en contraparte con la distribución Log-normal que presentó desviaciones mayores.

El algoritmo genético que se utilizó en este estudio, permitió estimar los parámetros de la DPG3 con mayor precisión, en comparación con el método de bisección. Esto permitió simular muestras para estimar luego los ICP usando los método no normales; se pudo constatar que el método de Clements generó resultados tanto de C_p como de C_{pks} más próximos a los valores objetivos de C_{pu} (caso contrario a los obtenidos a la DPG2), lo que permite decidir qué método es mejor para estimar los ICP en este caso. En comparación con una distribución Normal se nota la diferencia en ambos ICP, lo que puede conllevar al investigador o ingeniero de calidad a tomar decisiones erróneas respecto a un proceso productivo.

En síntesis, con esta investigación se concluye que sí es posible estimar ICP no normales usando la DPG2 y los métodos de percentiles de Clements y percentiles de Burr. Siendo el método de Burr el que mejores resultados produce. Por otra parte, se propuso el método exacto como otra alternativa para estimar estos ICP, lo que generó resultados mejores en comparación con los dos métodos antes mencionados; lo que lleva a pensar que también es posible usar este método exacto para estimar los ICP y cuyas respuestas son mejores sobre todo cuando la asimetría de las distribuciones es severa. Caso contrario ocurrió con la DPG3, con la cual también se puede estimar los ICP, pero los resultados más próximos a los valores targets fijados, en este caso los produjo el método de percentiles de Clements.

www.bdigital.ula.ve

c.c Reconocimiento

RECOMENDACIONES Y FUTUROS TRABAJOS

Si bien es cierto que el modelo de la distribución PG es algo complicado y los datos son algunas veces difíciles de conseguir, este si puede ser utilizado para estimar ICP. Como pudo observarse, las fórmulas del modelo de la DPG3 son extensas y complicadas al estimar sus parámetros, pero esto se puede resolver usando un algoritmo genético, el cual se presenta en la sección de anexos (anexo 6), Igualmente se pueden estimar los ICP usando los modelos de la DPG2 y DPG3. Sin embargo, cuando se logra construir el modelo de la DPG4 y se trata de estimar sus parámetros desconocidos bien sea por el método de máxima verosimilitud o de momentos, las fórmulas generadas se vuelven un tanto intratables. Queda abierta la posibilidad de utilizar este modelo como objetivo de estudio en la estimación de sus parámetros usando algoritmos genéticos, o como aplicación en la estimación de ICP.

www.bdigital.ula.ve

c.c Reconocimiento

BIBLIOGRAFÍA

Abd Elfattah, A. M., Elsherpieny, E. A. & Hussein, E. A. (2007). Parameters Estimations for New Generalized Pareto Distribution. Submitted.

Albing, M. (2006). Process capability analysis with focus on indices for one-sided specification limits. Lulea University of Technology.

Aguilar, J. y Rivas, F. (2001). Introducción a las Técnicas de Computación Inteligente. Editores Meritec, Mérida Venezuela.

Ahmad, S., Abdollahian, M. and Zeephongsekul, P. (2007). Process Capability Analisis for Non-Normal Quality Characteristics Using Gamma Distribution. Proceedings of the International Conference on Information Technology. 420-424. Arnald, B.B. and Press, S.J. (1983). Bayesian Inference for Pareto Populations. Journal of Econometrics, 21. 287-306.

Baxter, M. A. (1980). Minimum Variance Unbiased Estimation of the Parameter of the Pareto Distributions. Biometrika, 27. 133-138.

Bernardo J. & Irony T. (1996). A general multivariate Bayesian Process Capability index. The Statistician. 45. No 3. Pp 487-502.

Bernardo, J. & Smith, A. (1994). Bayesian Theory. Chi Chester: Wiley.

Bittanti, S., Lovera, M. and Moiraghi, L. (1998). Application of Non-normal Process Capability Indices to Semiconductor Quality Control. IEEE Transactions of Semiconductor Manufacturing. 11(2):296-303.

Bothe, D. (1999). Composite capability index for multiple product characteristics. Quality Engineering, 12(2), 253-258.

Box, G.E.P and Cox, D.R. (1964). An analysis of transformations. J.R. Statistics Soc. B, 26: 221-252.

Boyles, R.A. (1994). The Taguchi capability index. Journal of Quality Technology. 23: (1): 107-126.

- Breyfogle, F.W. (1999). *Implementing, Six Sigma*. Wiley: New York.
- Burr, IW. (1942). Cumulative frequency distribution. *Ann Math Stat.* 13:215-232.
- Burr, IW. (1967). The effect of non-normality on constants for \bar{X} and R charts. *Ind. Qual Control.* May: 563-569.
- Burr, IW. (1973). Parameters for a general system of distributions to match a grid of α_3 y α_4 . *Commun Stat.* 2:1-21.
- Castagliola, P. (1996). Evaluation of Non-normal Process Capability Indices using Burr's Distribution. *Qual. Engng.* **8**:587-593.
- Castagliola, P. y Castellanos J, V G (2005) Capability indices dedicated to the two quality characteristic cases. *Qual Tech Quantitative Management* 2:201-220
- Chan, L. K., Cheng, S. W. y Spiring, F. A. (1991). A multivariate measured of process capability: Cpm. *Internat. J. Model. Simul.* **11**, 1-6.
- Chang, Y.S. Bai, D.S. (2002). Process Capability indices for skewed populations. *Quality and Reliability Engineering International.* **17**: 397-406.
- Chatterjee, S. and Laudato, M. (1997). *Genetics Algorithms in Statistics*. Common Statistics.
- Cheng, S & Spiring, F. (1989). A new measured of process capability: C_{pm}. *Journal of Quality Technology.* 20. 162-175.
- Choi, L.S. and Bai, D.S. (1995). X-bar and R control charts for skewed populations. *J. Qual. Technol.*, **27**: 120-131.
- Choi, L.S. and Bai, D.S. (1997). Process capability indices for skewed populations. *Proc. 20th. Conf. on Computer and Industrial Engineering.* Pp. 1211-1214.
- Choobineh, F. and Branting, D. (1986). A Simple Approximations for Semivariance. *Eur. J. Res.*, **27**:364-370.
- Choobineh, F. and Ballard, J.L. (1987). Control-limits of QC charts for Skewed distributions using weighted-variance. *IEEE Trans. Reliab.* **REL-36**. 473-477.

Chow, C.Y., Cheng, P.H. (1997). Ranges control chart for non-normal data. *J. Chinese Ints Ind Eng.* **14** (4): 401-409.

Chow, V. and Owen, D. (1998). On the Distribution of the Estimated Process Capability Indices. *Comm. Statist. Theory and Methods.* **18.** 4549-4560.

Chou, Y.M., Polansky, A.M. and Mason, R.L. (1998). Transforming Non-normal data to normality in statistical Process Control. *Journal of Quality Technology.* **30**:133-141.

Choulakian, V. Y Stephens, M. A. (2001). Goodness of fit test for the Generalized Pareto Distribution. *Technometrics.* Vol 43. No. 4. 478-484.

Clements, J.A. (1989). Process Capability Calculations for Non-Normal Distribution, *Quality Progress,* **24** (9), 95-100.

Cook, W. L. & Mumme, D.C. (1981). Estimation of Pareto parameter by Numerical methods. In: *Statistical Distributions in Scientifics Work*, ed. C. Taillie et al. **5** 127-132.

Czarski, A. (2008). Estimation of process capability indices in case of distribution unlike the normal one. *Archives of materials science and Engineering.* **34**: 1: 39-42.

Davinson, A.(1984). Modeling excesses over high Threshold, with Applications. In: *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, 461-482. Reidel, Dordrecht, the Netherlands.

Davinson, A. C. (1984b). A Statistical Model for Contaminations due two long-range atmospheric transport of radionuclides. PhD thesis. Department of Mathematics. Imperial College of Science and Thechnology, London, UK.

Davinson, A. C. & Smith, R. L. (1990). Models for Exceedances Over High Thresholds. *J. Roy. Statis. Soc. B.* **52(3)**, 393-442.

DeJong, K.(1975). An analysis of the Behavior of a Class of Genetic Adaptative Systems, Ph.D. Thesis, University of Michigan.

Deleryd, M. Vännman, K. (1999). Process capability plots – a quality improvement tool. *Qual. Reliab. Engng. Inst.* **15** 213-227.

- Ding, J. (2004). A Method of Estimating the Process Capability Index from the First Four Moments of Non-normal Data. *Quality and Reliability Engineering International*. **20**: 787-805.
- DuMouchel, W. (1983). Estimating the Stable Index α in Order to Measure Tail Thickness. *Ann. Statist.* **11**, 1019-1036.
- Efron, B. & Tibshirani, R. (1986). Bootstrap Methods for Estimating Errors Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1. 54-77.
- Elderton, W. P. y Johnson N. (1969) *Systems of frequency curves* Cambridge, UK: Cambridge Univ Press
- Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal events for insurance and Finance. Applications of mathematics.* Springer.
- English J.R. and Taylor, G.D. (1993). Process capability analysis- a robustness study. *Int. J. Prod. Res.*, **31**:1621-1635.
- Farnum, N.R. (1996). Using Johnson curves to describe non-normal data. *Quality Engineering*. **9**:329-336.
- Fermín J. et al. (2009). Control Estadístico de Procesos multivariantes en la industria alimentaria: Implementación a través del estadístico T^2 -Hotelling. *Revista Agroalimentaria*. Vol. 15. Nro 28. Enero-Junio. 91-105.
- Flaig, J.J. (1996). A new approach to process capability analysis. *Quality Engineering*, **9**:205-211.
- Franklin, L.A. And Wasserman, G. (1992). Bootstrap confidence interval estimates of C_{pk} : An introduction. *Communications in Statistics Series B*. **20**:231-242.
- Franklin, L.A. and Wasserman, G. (1998). Bootstrap confidence limits for capability indices. *Journal of Quality Technology*. **24**:196-210.

- Gen, M. and Cheng, R. (1996). Genetic Algorithms and Engineering Desing. Wiley, New York.
- Goldberg, D. (1989). Genetic Algorithms In Search, Optimization and Machine Learning (Reading, MA: Addison-Wesley).
- Graybill, F. Theory and Application of the Linear Model. Duxbury Press. Unite States of America.
- Gruska, G.F., Mirkhani, K. and Lamberson, L.R. (1989). Non-normal Data Analysis. (St. Clair Shores, MI:Appllied Computer Solutions, Inc).
- Gunter, B.H. (1989). The Use and abuse of C_{pk} : Parts 1-4. Part 3. *Quality Progress*. May, 79-80.
- Gulsen, M.et al. (1995). A genetics Algorithm Approach to Curve Fitting. Int. J. Prod. Res. Vol. 33, No. 7. 1911-1923.
- Hahn, G.J. and Shapiro, S. (1997). Statistical Models in Engineering. Wiley. New York.
- Han, C. (2006). Effect of Testing Normality on Estimating Process Capability Indices. *Quality Engineering*, 18:391-395.
- Heuvel, E y Ion, R (2003) Capability Indices and the proportion of non conformin items *Quality engineering* vol. 15, pp 427-439.
- Hill, L.D., Hill, R. and Holder, L. (1976). Fitting Johnson Curves by moments algorithm AS99. *Appl Statist.* **25**: 180-189.
- Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- Hosking, J. and Wallis, J.(1987). Parameter and Quantiles Estimation for the Generalized Pareto Distribution. *Technometrics*, Vol. 29. N. 3. 339-349.
- Hubele, N.F., Shahriari, H. Y Cheng, C.S. (1991). A bivariate process capability vector. *Statistical Process in Manufacturing* (Edited by J.B. Keats and D.C. Montgomery), 299-310. Marcel Dekker, New York.
- Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149-176.

- Johnson, N. et al. (1994). Continuous Univariate Distributions. Vol. 1. New York: John Wiley and Sons.
- Johnson, N. et al. (1992). Flexible Process Capability Indices. Pakistan Journal of Statistics, **10**:23-31.
- Juran, J. M. and Gryna, F.M. (1980). Quality Planning and Analysis. Mc.Graw-Hill. New York. New York.
- Kane, V.E. (1986). Process Capability Indices. Journal of Quality Technology, **18**. 41-52.
- Kendall, M y Stuart A. (1969) The Advanced theory of statistics. London, U. K Charles Griffin.
- Kotz, S. and Johnson N.L. (1993). Process Capability Indices. Chapman and Hall. London.
- Kotz, S. and Johnson N.L. (2002). Process capability indices- a review (1999-20000). Journal of quality Technology. **34**: (1): 2-19.
- Kotz, S. and Lovelace, C.R. (1998). Process Capability indices in Theory and Practice. Arnold: London.
- Liao, S. et al. (2002) Capability evaluation for process of the larger the better type for non-normal populations Adv. Appl Stat 2: 189-198.
- Littig, S. et al. (1992). Process Capability measurements for a bivariate characteristics over an elliptical tolerance zone. Technical Report No. 92-42, Department of Industrial and Operations Engineering, The University of Michigan.
- Liu, P.S. and Chen, F.L. (2006). Process Capability Analysis on Non-normal Process Data Using the Burr XII Distribution. The International Journal of Advanced Manufacturing Technology. **27**: 975-984.
- Lu, M. W. and Rudy, R.J. (2002). Discussion on 'Process Capability Indices –a review, 1992-2000'. Journal of Quality Technology. **34**:38-39.
- MatLab, Versión 7. (1992). The Language of Technical Computing. Copyright 1984-2004. The MathWorks Inc.

Mathews, J.H.(1992). Numerical Methods for Mathematics, Science and Engineering. Prentice Hall International, Inc., London, 642 p.

Montgomery, D.C. (2005). Introduction to Statistical Quality Control. 2da edition. John Wiley and Sons.

Munechika, M. (1986). Evaluation of Process Capability for skew distributions paper presented to proceedings 30th EOQ Conference Stockholm

Munechika, M. (1992). Studies on process capability in machining processes. Reports of statistical Applied research JUSE 39: 14-29.

Nahar, P.C., Hubele, N.F. and Zimmer, L.S. (2001). Assessment of capability index to skewness. Quality and Reliability Engineering International. **17**:233-241.

Öztürkler, H. and Altan S. (2008). A Genetic Algorithm Approach to parameter Estimation in Nonlinear Econometric Model. Dumlupınar Üniversitesi Sosyal Bilimler Dergisi. Pp: 67-76.

Pal, S. (1999). Performance evaluation of a bivariate normal process. Quality Engineering, 11(3), 379-386.

Pal, S. (2005). Evaluation of nonnormal process Capability Indices using Generalized Lambda Distribution. Quality Engineering. 17:77-85.

Pearn, W.L. and Chen, K.S. (1997). Capability Indices for Non-normal Distributions with an applications in electrolytic Capacitor Manufacturing. Microelectronics and Reliability, **37**:1853-1858.

Pearn, W.L. and Kotz, S. (1994). Application of Clements' method for calculating second and third generation process capability indices for non-normal Pearsonian populations. Quality Progress. **7**(1):139-145.

Pearn, W.L., Kotz, S. Y Jhonson, N.L. (1992). Distributional and Inferential properties of process capability indices. Journal of Quality Technology, 24, 216-231.

- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *Ann. Statist.* 3. 119-131.
- Polak, E. (1997). *Optimization Algorithms and Consistent Approximations*. Springer, 767 p.
- Polansky, A.M. (2000). An Algorithm for Computing Non-parametric Capability Estimate. *Journal of Quality Technology*. **32**:284-289.
- Polansky, A.M. (2001). A Smooth Nonparametric Approach to Multivariate Process Capability. *American Statistical Association and the American Society for Quality Technometrics*. Vol 43, NO. 2. 199-211.
- Polansky, A.M. (1998). A Smooth Nonparametric Approach to Process Capability. *Quality and Reliability Engineering International*. **14**:43-48.
- Polansky, A. et al. (1998). Estimating process capability indices for a truncated distribution. *Quality Engineering*, 11 (2): 257-265.
- Polansky, A. et al. (1999). A Algorithm for Fitting Johnson Transformation to Non-normal Data. *Journal of Quality Technology*. **31**: 345-350.
- Press, W. et al. (1992). *Numerical Recipes in Fortran*, Cambridge University Press, Cambridge, New York, 963 p.
- Pyzdek T. (1992). Process capability using personal computers. *Quality Engineering*; **4**:419-440.
- Quandt, R.E. (1966). Old a New Methods of Estimation of the Pareto Distribution. *Biometrika* **10**. 55-82.
- R Development Core Team.(2008). *R: A Language and Environment for Statistical*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>
- Rheinboldt, W.C. 1998. *Methods for Solving Systems of Nonlinear Equations*, SIAM, PA, USA. 148 p.

Rivas, D. et al. (2006). Aplicación de los Algoritmos Genéticos para Estimar los Parámetros en un Modelo de Regresión de Cox. *Economía*, XXXI, 22. Pp. 57-74.

Rivera, L. A., Hubele, N.F. y Lawrence, F.D. (1995). C_{pk} index estimation using data transformation. *Compt. Ind. Engng*, **29**, 55-58.

Rodríguez, R. N. (1992). Recent Developments in Process Capability Analysis. *Journal of Quality Technology*, **24**:176-186.

Rosbjerg, D. Madsen, H. & Rasmussen, P.F. (1992). Prediction in Partial Duration Series with Generalized Pareto-distributed Exceedances. *Wat Resour. Res.* **28(11)**. 3001-3010.

Saksena, S. K. and Johnson, A. M. (1984). Best Unbiased Estimators for the Parameters of a two-parameter Pareto Distribution. *Biometrika*, **31**. 77-83.

Sarkar, A. and Pal, S. (1997). Estimation of process capability index for Concentricity. *Quality Engineering*, **9**: 665-671.

Shore, H. (1998). A new approach to analyzing non-normal quality data with application to process capability analysis. *International Journal of Production Research*. **36**: 1917-1933

Singh, V. P. & Guo H. (1995). Parameter Estimation for 3-parameter Generalized Pareto Distributions by the Principle of Maximum Entropy (POME). *Hydrological Sciences. J. Des Science Hydrollogiques*, **40**, 2. 165-181.

Singpurwalla, N. (1998). The stochastic control of process capability indices *Sociedad de estadística e Investigación Operativa Test*, vol 7, nº 1, pp 1-74.

Smith, R. L. (1984). Threshold Methods for Sample Extremes. In: *Statistics Extremes and Applications* ed. J. Trago de Oliveira, 621-638. Reidel, Dordrecht. The Netherlands.

Smith, R. L. (1987). Estimating Tails Of Probability Distributions. *Ann. Statist.*, **15**. 1174-1207.

Smith, J.A. (1991). Estimating the Upper Tail of Flood Frequency Distributions. *Wat Resour. Res.* **23(8)**. 1657-1666.

Spiring, F. et al. (2003). A bibliography of process capability papers. *Quality and Reliability Engineering International* 19, 445-460.

Somerville, S. E. and Montgomery, D.C. (1997). Process Capability Indices and Non-normal Distributions. *Qual. Engng.*, 9: 305-316.

S-PLUS Professional. 2001. Release 2. Copyright (c). Insightful Corp.

Taam, W. et al. (1993). A note on multivariate capability indices. *Journal of Applied Statistics*, 20 (3), 339-351.

Tang, L.C. & Than S.E. (1999). Computing Process Capability Indices for Non-normal Data: a review and comparative study. *Quality and Reliability Engineering International*. 15: 339-353.

Telmoudi, R. (2005). A multi-stream process capability assessment using a nonconformity ratio based desirability function. University of Dortmund.

Vänman, K. (1976). Estimators Based of Order Statistics from a Pareto Distributions. *Journal of the American Statistical Associations*. Vol. 71. NO. 355. 704-708.

Vänman, K. (1995). A Unified Approach to Capability Indices. *Statis. Sinica*, 5: 805-820.

Vännman K. and Albing M. (2007). Process capability indices for one-sided specification intervals and skewed distribution. *Quality and Reliability Engineering International*. **23**: 755-765.

Van Montfort, M.A. & Witter, J.V. (1985). Testing Exponentially Against Generalized Pareto Distributions. *J. Hydrol.* 78. 305-315.

Van Montfort, M.A. & Witter, J.V. (1986). The Generalized Pareto Distributions Applied to Rainfall depths. *Hydrol. Sci. J.* 31(2). 151-162.

Van Montfort, M.A. & Otten, A. (1991). The Firts and the Seconds e of the Extreme Value Distribution. EV1. *Stochastic Hydrol. Hidraul*, 5, 69-76.

Wang, F.K. y Chen, J.C. (1998). Capability index using principal components analysis. *Quality Engineering*, (11)1, 21-27.

Wang, F. K. y Du, T. (2000). Using principal component analysis in process performance for multivariate data. *Omega: The international Journal of Management Science*, 28, 185-194.

Wright, P.A. (1995). A Process Capability Index Sensitive to Skewness. *J. Statis. Comput. Simul.*, **52**: 195-203.

Wu, H-H, Swain, J. J. Farrington, P.A. and Messimer, S. L (1999). A Weighted Variance Capability Index for General Non-normal Process. *Quality and Reliability Engineering International*, **15**: 397-402.

Wu, H. and Swain, J. J. (2001). A Monte Carlo Comparison of Capability Indices when Process Non-Normally Distributed. *Quality and Reliability Engineering International*, **17**: 219-231.

Wu H.H. et al. (1998). Discussions of the Clements based process capability indices In. *Proceeding of the 1998. CHE National Conference*. Pp 561-566.

Yerniay, M. (1999). Taguchi Deneý Tasarimi Problemlerine Genetik Algoritma Yaklasimi.

Zimmer, W.J and Burr IW. (1963). Variables sampling plans based on non-normal population. *Ind. Qual. Control*. July:18-36.

ANEXOS

www.bdigital.ula.ve

c.c Reconocimiento

ANEXO 1: Permite estimar los ICP usando la metodología de Liu-Chen usando Software R.

1.1. Simulación de la Distribución Weibull. Método para estimar los ICP mediante percentiles de Burr y percentiles de Clements (Usando R)

#Inicio: simulación de la muestral con los parámetros conocidos

```
x.wei<-rweibull(100,shape=1.0,scale=1.2)
```

```
media.wei<-mean(x.wei)
```

```
media.wei
```

```
s.wei<-sqrt(var(x.wei))
```

```
s.wei
```

```
t.wei<-median(x.wei)
```

```
t.wei
```

```
asim.wei<-skewness(x.wei)
```

```
asim.wei
```

```
kurto<-kurtosis(x.wei)
```

```
kurto
```

```
quantile(x.wei,c(0.99865,0.50,0.00135))
```

```
#LES<-Cpu(X99865-X50)+X50
```

```
a<-n-2
```

```
b<-sqrt(n*(n-1))
```

```
cociente<-a/b
```

```
cociente
```

```
alfa3<-cociente*asim.wei
```

```
alfa3
```

```
a1<-(n-2)*(n-3)
```

```
b1<-(n+1)*(n-1)
```

```
cociente1<-a1/b1
```

```
kurto.wei<-kurtosis(x.wei)
```

```

cociente2<-(n-1)/n+1
Zstar00135<- -0.708
Zstar99865<-5.15423
Zstar50<-0.484
#hallar X00135 conocido Zstar00135
X00135.wei<-media.wei + Zstar00135*s.wei
X00135.wei
X50.wei<-media.wei + Zstar50*s.wei
X50.wei
X99865.wei<-media.wei + Zstar99865*s.wei
X99865.wei
LES.wei<-7.4142
Cpu.wei<-(LES.wei-X50.wei)/(X99865.wei- X50.wei)
Cpu.wei

```

1.2.Simulación de la Distribución Beta. Método para estimar los ICP mediante percentiles de Burr y percentiles de Cléments (Usando R)

#Inicio: simulación de la muestral con los parámetros conocidos

```

set.seed(1001)
n<-100
x.beta<-rbeta(100,shape1=4.4,shape2=3.3)
media.beta<-mean(x.beta)
media.beta
s.beta<-sqrt(var(x.beta))
s.beta
t.beta<-median(x.beta)
t.beta
asim.beta<-skewness(x.beta)
asim.beta

```

```

kurto<-kurtosis(x.beta)
kurto
quantile(x.beta,c(0.99865,0.50,0.00135))
#LES<-Cpu(X99865-X50)+X50
#LES<-Cpu(X99865-X50)+X50
a<-n-2
b<-sqrt(n*(n-1))
cociente<-a/b
cociente
alfa3<-cociente*asim.beta
alfa3
a1<-(n-2)*(n-3)
b1<-(n+1)*(n-1)
cociente1<-a1/b1
kurto.beta<-kurtosis(x.beta)
cociente2<-(n-1)/n+1
#hallar c y k conocidos alfa3 y alfa4
alfa4<-cociente1*kurto.beta + 3*cociente2
alfa4
alfa3
Zstar00135<- -3.642
Zstar99865<-3.659
Zstar50<-0.001
#hallar X00135 conocido Zstar00135
X00135.beta<-media.beta + Zstar00135*s.beta
X00135.beta
X50.beta<-media.beta + Zstar50*s.beta

```

```

X50.beta
X99865.beta<-media.beta + Zstar99865*s.beta
X99865.beta
LES.beta<-0.94229
Cpu.beta<-(LES.beta-X50.beta)/(X99865.beta- X50.beta)
Cpu.beta

```

1.3.Simulación de la Distribución Weibull. Método para estimar los ICP mediante percentiles de Burr y percentiles de Clements (Usando S-plus)

#Inicio: simulación de la muestral con los parámetros conocidos

```
set.seed(3546)
```

```
n=100
```

```
x<-matrix(rweibull(n,1.0,1.2))
```

```
hist(x)
```

```
xhart<-mean(x)
```

```
s<-sqrt(var(x))
```

```
sk<-skewness(x)
```

```
k<-kurtosis(x)
```

```
#percentiles de la distribución
```

```
Lp<-quantile(x,0.00135)
```

```
M<-quantile(x,0.5)
```

```
Up<-quantile(x,0.99865)
```

```
#momentos estandarizados de asimetria y kurtosis
```

```
alpha3<-((n-2)*sk/sqrt(n*(n-1)))
```

```
alpha4<-(((n-2)*(n-3)*k)/(n^2-1))+ 3*((n-1)/(n+1))
```

```
medidas<-c(xhart,s,sk,k,Lp,M,Up)
```

```
medidas
```

```
alpha3
```

```
alpha4
```

```
#Para generar el Ut
```

```
Cpu=2.0
```

```
Ut<-Cpu*(Up-M)+M
```

```
Ut
```

```
Lt=0.1
```

```
#Los índices de Clements
```

```
CCp<-((Ut-Lt)/(Up-Lp))
```

```
CCpu<-((Ut-M)/(Up-M))
```

```
CCpl<-((M-Lt)/(M-Lp))
```

```
CCpk<-min(CCpu,CCpl)
```

```
CCp
```

```
CCpk
```

```
#Para construir: Lp, M, Up de Burr
```

```
BZ0135=-1.671
```

```
BZ5=-0.157
```

```
BZ99865=4.390
```

```
#procedimiento para los Z con Burr
```

```
BLp<-xhart+ s*BZ0135
```

```
BM<-xhart+ s*BZ5
```

```
BUp<-xhart+s*BZ99865
```

```
BLp
```

```
BM
```

```
BUp
```

```
#Para el Ut otra vez
```

```
Cpu=2.0
```

```
Ut<-Cpu*(Up-M)+M
```

```
Ut
```

```
Lt=0.1
```

```
#Se calculan los ICP de Burr
```

```
BCp<-(Ut-Lt)/(BUp-BLp)
```

```
BCpu<-(Ut-BM)/(BUp-BM)
```

```
BCpl<-(BM-Lt)/(BM-BLp)
```

```
BCpk<-min(BCpu,BCpl)
```

```
BCp
```

```
BCpk
```

www.bdigital.ula.ve

1.4. Simulación de la Distribución Beta. Método para estimar los ICP mediante percentiles de Burr y percentiles de Clements (Usando S-plus)

```
#Inicio: Generación de la muestra
```

```
set.seed(35461)
```

```
n=100
```

```
x<-matrix(rbeta (n,4.4,13.3))
```

```
hist(x)
```

```
xhart<-mean(x)
```

```
s<-sqrt(var(x))
```

```
sk<-skewness(x)
```

```
k<-kurtosis(x)
```

```
#percentiles de la distribución
```

```
X000135<-quantile(x,0.00135)
```

```
X05<-quantile(x,0.5)
```

```
X099865<-quantile(x,0.99865)
```

```
#momentos estandarizados de asimetria y kurtosis
```

```
alpha3<-(n-2)*sk/sqrt(n*(n-1))
```

```
alpha4<-((n-2)*(n-3)*k/(n^2-1))+ 3*((n-1)/(n+1))
```

```
medidas<-c(xhart,s,sk,k)
```

```
#Para generar el Ut
```

```

Cpu=2.0
Ut<-Cpu*(X099865-X05)+X05
medidas
X000135
X05
X099865
alpha3
alpha4
Ut
#Para construir los LP(con Z0.0135) M(con Z0.5) Up(con Z0.99865)buscados en la tabla
BZ0135=-0.596
BZ5=-0.529
BZ99865=4.291
#procedimiento para los Z con Burr
BLp<-xhart+ s*BZ0135
BM<-xhart+ s*BZ5
BUp<-xhart+s*BZ99865
BLp
BM
BUp
#Un valor Lt dado
Lt=0.1
#Se calculan los ICP de Burr
BCp<-(Ut-Lt)/(BUp-BLp)
BCpu<-(Ut-BM)/(BUp-BM)
BCpl<-(BM-Lt)/(BM-BLp)
BCpk<-min(BCpl,BCpu)
BCp
BCpl
BCpu
BCpk
#Para construir los LP(con Z0.0135) M(con Z0.5) Up(con Z0.99865)buscados en la tabla
CZ0135=-0.618
CZ5=-0.513
CZ99865=4.072
CLp<-xhart+s*CZ0135
CM<-xhart+ s*CZ5
CUp<-xhart+s*CZ99865
CLp
CM
CUp
#Se calculan los ICP de Clements
Lt=0.1
CCp<-(Ut-Lt)/(CUp-CLp)
CCpu<-(Ut-CM)/(CUp-CM)
CCpl<-(CM-Lt)/(CM-CLp)

```

```

CCpk<-min(CCpu,CCpl)
CCp
CCpl
CCpu
CCpk

```

ANEXO 2: Construcción de la DPG2 conocidos sus parámetros, simulación de las muestras y estimación de los ICP no normales (Método de percentiles de Burr y Clements mediante el software S-plus

#Inicio: Construcción de la Distribución Pareto Generalizada con 2 parámetros (DPG2)

```

dpareto<-function(x,k=-0.2,a=1) (1/a)*(1-k*x/a)^((1-k)/k)
ppareto<-function(x,k=-0.2,a=1) (x>a)*(1-(1-(k*x/a)))^(1/k)
qpareto<-function(u,k=-0.2,a=1) a/(1-u)^(1/k)
rpareto<-function(n,k=-0.2,a=1) qpareto(runif(n),k,a)
x<-seq(1,50,len=100)
plot(x,dpareto(x))
plot(x,ppareto(x))
plot(x,rpareto(x))
u<-seq(0.02,0.9,len=100)
plot(u, qpareto(u))
#para generar cualquier muestra PG2
y<-rpareto(100)
y
hist(y)

```

#Para generar una muestra con una semilla DPG2

```

set.seed(3876343)
n=100
PG<-matrix(rpareto(n,-0.2,1))
hist(PG)

```

#Buscar las propiedades de la DPG2

```

k=-0.2
a=1
n=100
media<-a/(1+k)
Var<-a^2/(((1+k)^2)*(1+2*k))
desv<-sqrt(Var)
sk<-2*(1-k)*((1+2*k)^0.5)/(1+3*k)
kurt<-(3*(1+2*k)*(3-k+2*(k^2)))/((1+3*k)*(1+4*k))-3
sk
kurt

```

#Quantiles

```

X0135<-quantile(y,0.00135)

```

```

X05<-quantile(y,0.5)
X99865<-quantile(y,0.99865)
medidas<-c(media,desv,sk,kurt,X0135,X05,X99865)
medidas

#Los valores de alpha3 y alpha 4
alpha3<-((n-2)*sk)/sqrt(n*(n-1))
alpha4<-(((n-2)*(n-3)*kurt)/(n^2-1))+ 3*((n-1)/(n+1))
alpha3
alpha4

#Para generar el Ut
Cpu=1.0
Ut<-Cpu*(X99865-X05)+X05
Ut
#Para construir los LP(con Z0.0135) M(con Z0.5) Up(con Z0.99865)buscados en la tabla
BZ0135=-0.596
BZ5=-0.529
BZ99865=4.291

#procedimiento para los Z con Burr
BLp<-xhart+ s*BZ0135
BM<-xhart+ s*BZ5
BUp<-xhart+s*BZ99865
BLp
BM
BUp

#Un valor Lt dado
Lt=0.1

#Se calculan los ICP de Burr
BCp<-(Ut-Lt)/(BUp-BLp)
BCpu<-(Ut-BM)/(BUp-BM)
BCpl<-(BM-Lt)/(BM-BLp)
BCpk<-min(BCpl,BCpu)
BCp
BCpl
BCpu
BCpk

#Para construir los LP(con Z0.0135) M(con Z0.5) Up(con Z0.99865)buscados en la tabla
CZ0135=-0.618
CZ5=-0.513
CZ99865=4.072
CLp<-xhart+s*CZ0135
CM<-xhart+ s*CZ5

```

```
CUp<-xhart+s*CZ99865
CLp
CM
CUp
```

```
#Se calculan los ICP de Clements
```

```
Lt=0.1
CCp<-(Ut-Lt)/(CUp-CLp)
CCpu<-(Ut-CM)/(CUp-CM)
CCpl<-(CM-Lt)/(CM-CLp)
CCpk<-min(CCpu,CCpl)
CCp
CCpl
CCpu
CCpk
```

ANEXO 3: Estimación de los ICP dada una muestra de datos reales mediante la metodología de Liu-Chen (usando S-plus)

```
x<-
  c(5,5,25,27,15,35,33,17,15,12,1,3,4,16,26,35,10,54,6,18,4,3,2,8,15,17,33,26,33,22,17,
    2,1,3,21,17,27,16,27,18,14,2,1,8,8,8,8,5,4,1,4,7,5,10,69,56,73,61,49,38,25,17,9,27,36,
    20,34,24,13,4,4,21,42,56,42,31,17,4,28,1,1,13,57,73,73,108,41,3,24,8,1,34,14,45,27,7,
    .....10,11,13)
summary(x)
hist(x)
xhart<-mean(x)
s<-sqrt(var(x))
sk<-skewness(x)
k<-kurtosis(x)
#percentiles de la distribución
Lp<-quantile(x,0.00135)
M<-quantile(x,0.5)
Up<-quantile(x,0.99865)
#momentos estandarizados de asimetria y kurtosis
alpha3<-(n-2)*sk/sqrt(n*(n-1))
alpha4<-((n-2)*(n-3)*k/(n^2-1))+ 3*((n-1)/(n+1))
medidas<-c(xhart,s,sk,k,Lp,M,Up)
#Para generar el Ut
Cpu=2.0
Ut<-Cpu*(Up-M)+M
medidas
alpha3
alpha4
Ut
#Para construir los LP(con Z0.0135) M(con Z0.5) Up(con Z0.99865)buscados en la tabla
```

```
BZ0135=-0.828
BZ5=-0.435
BZ99865=4.452
```

```
#procedimiento para los Z con Burr
```

```
BLp<-xhart+ s*BZ0135
BM<-xhart+ s*BZ5
BUp<-xhart+s*BZ99865
BLp
BM
BUp
#Un valor Lt dado
Lt=0.1
#Se calculan los ICP de Burr
BCp<-(Ut-Lt)/(BUp-BLp)
BCpu<-(Ut-BM)/(BUp-BM)
BCpl<-(BM-Lt)/(BM-BLp)
BCpk<-min(BCpu,BCpl)
```

```
#Se calculan los ICP de Clements
```

```
Lt=0.1
CCp<-(Ut-Lt)/(Up-Lp)
CCpu<-(Ut-M)/(Up-M)
CCpl<-(M-Lt)/(M-Lp)
CCpk<-min(CCpu,CCpl)
CCp
CCpk
BCp
BCpk
```

ANEXO 4: Método Bootstrap para generar los intervalos bootstrap y cubrimientos para una muestra real usando software S-plus

```
y<-
  c(5,5,25,27,15,35,33,17,15,12,1,3,4,16,26,35,10,54,6,18,4,3,2,8,15,17,33,26,33,22,17,
    2,1,3,21,17,27,16,27,18,14,2,1,8,8,8,5,4,1,4,7,5,10,69,56,73,61,49,38,25,17,9,27,36,
    20,34,24,13,4,4,21,42,56,42,31,17,4,28,1,1,13,57,73,73,108,41,3,24,8,1,34,14,45,27,7,
    .... ,10,11,13)
summary(y)
hist(y, xlab="Precipitación con tiempos de 1 hora (mm/hora)", ylab="Frecuencias")
#PASO 1: ESTIMAR MEDIA, DESV, ASIM, KURT, PERCENTILES.
#percentiles de la distribución
X000135<-quantile(y,0.00135)
X05<-quantile(y,0.5)
X099865<-quantile(y,0.99865)
n<-length(y)
```

```

n
xhart<-mean(y)
s<-sqrt(var(y))
sk<-skewness(y)
k<-kurtosis(y)

#PASO 2: CON LA ASIM Y KURT SE CALCULAN ALPHA3 Y ALPHA4 PARA
      DETERMINAR LOS BZ* Y Z*
#momentos estandarizados de asimetria y kurtosis
#momentos estandarizados de asimetria y kurtosis
alpha3<-(n-2)*sk/sqrt(n*(n-1))
alp3<-(n-2)*sk/sqrt(n*(n-1))
alp4<-((n-2)*(n-3)*k/(n^2-1))+(3*(n-1)/(n+1))
medidas<-c(xhart,s,sk,k,X000135,X05,X099865,alp3,alp4)
medidas

#Para generar el Ut
Cpu=1.5
Ut<-Cpu*(X099865-X05)+X05
Ut
Lp<-quantile(y,0.00135)
M<-quantile(y,0.5)
Up<-quantile(y,0.99865)
Lp
M
Up

#PASO 3: BUSCAR LOS VALORES TABULADOS
#Para construir los LP(con BZ0.0135) M(con BZ0.5) Up(con BZ0.99865)buscados en la
      tabla
BZ0135=-0.650
BZ5=-0.516
BZ99865=4.877
#Para construir los LP(con Z0.0135) M(con Z0.5) Up(con Z0.99865)buscados en la tabla
CZ0135=-0.758
CZ5=-0.416
CZ99865=5.012
#PASO 4: CÁLCULO DE LOS PERCENTILES ESTANDARIZADOS PARA BURR Y
      PARA CLEMENTS

#procedimiento para los percentiles de Burr
BLp<-function(y)(mean(y)+ sqrt(var(y))*BZ0135)
BM<-function(y)(mean(y)+ sqrt(var(y))*BZ5)
BUp<-function(y)(mean(y)+ sqrt(var(y))*BZ99865)
BLp(y)
BM(y)
BUp(y)
#procedimiento para los percentiles de Clements

```

```
CLp<-function(y)(mean(y)+ sqrt(var(y))*CZ0135)
CM<-function(y)(mean(y)+ sqrt(var(y))*CZ5)
CUp<-function(y)(mean(y)+ sqrt(var(y))*CZ99865)
CLp(y)
CM(y)
CUp(y)
```

#PASO 5: CÁLCULO DE LOS ÍNDICES NO NORMALES USANDO LAS FÓRMULAS RESPECTIVAS.

```
Lt=1
#ICP de Burr
BCp<-function(y) (Ut-Lt)/(BUp(y)-BLp(y))
BCpu<-function(y)(Ut-BM(y))/(BUp(y)-BM(y))
BCpl<-function(y) (BM(y)-Lt)/(BM(y)-BLp(y))
BCpk<-function(y) min(BCpl(y),BCpu(y))
#ICP de Clements
CCp<-function(y) (Ut-Lt)/(CUp(y)-CLp(y))
CCpu<-function(y)(Ut-CM(y))/(CUp(y)-CM(y))
CCpl<-function(y)(CM(y)-Lt)/(CM(y)-CLp(y))
CCpk<-function(y) min(CCpu(y),CCpl(y))
```

#Resultados de Índices de Burr y Clements

```
BCp(y)
BCpk(y)
CCp(y)
CCpk(y)
```

#Índice Cpk con distribución normal

```
cp<-function(y) (Ut-Lt)/(6*sqrt(var(y)))
cpl<-function(y)(mean(y)-Lt)/(3*sqrt(var(y)))
cpu<-function(y)(Ut-mean(y))/(3*sqrt(var(y)))
cpk<-function(y) min(cpu(y),cpl(y))
cp(y)
cpl(y)
cpu(y)
cpk(y)
```

#Generando la muestra bootstrap para los intervalos

```
resamples<-lapply(1:1000, function(i) sample(y,replace=T))
r.BCpk<-sapply(resamples,BCpk)
r.BCpk
mean(r.BCpk)
desvr<-sqrt(var(r.BCpk))
desvr
asim<- skewness(r.BCpk)
asim
#ordenando la muestra obtenida
```

```

rr.BCpk<-sort(r.BCpk)
rr.BCpk

#Generando un intervalo percentil bootstrap.
b1<-c(rr.BCpk[25],rr.BCpk[975])
b1
#Generando un intervalo estándar bootstrap
h1<-c(mean(rr.BCpk)-1.96*sqrt(var(rr.BCpk)),mean(rr.BCpk)+1.96*sqrt(var(rr.BCpk)))
h1
#Percentil bootstrap corregido-insesgado.
zo<-qnorm(sum(rr.BCpk<BCpk(y))/1000)
P11<-pnorm(2*zo-1.96)
Pu1<-pnorm(2*zo+1.96)
a11<-P11*1000
a21<-Pu1*1000
a11
a21
m1<-c(rr.BCpk[a11],rr.BCpk[a21])
m1
#Buscando el cubrimiento del intervalo percentil bootstrap
mu=mean(b1)
mu
desvb1<-sqrt(var(b1))
desvb1
n.draw=100
n=100
draws=matrix(rnorm(n.draw*n,mu,desvb1),n)

#comandos para extraer los intervalos de confianza de 95% (conf.int)
get.conf.int=function(b1) t.test(b1)$conf.int
conf.intb=apply(draws,2,get.conf.int)
conf.intb
#promedio de los intervalos b1
Mb=mean(conf.intb)
db1=mean(sqrt(var(conf.intb)))
Mb
db1
sum(conf.intb[1,]<=Mb&conf.intb[2,]>=Mb)
plot(range(conf.intb),c(0,1+n.draw),type="n",xlab="intervalos percentil bootstrap",ylab="muestra")
for(i in 1:n.draw) lines(conf.intb[,i],rep(i,2),lwd=0.7)
  abline(v=mu, lwd=0.7, lty=0.70)
#Buscando el cubrimiento del intervalo estándar bootstrap
muh=mean(h1)
muh
desvh1<-sqrt(var(h1))
desvh1

```

```

n.draw=100
n=100
draws=matrix(rnorm(n.draw*n,muh,desvh1),n)
draws

#comandos para extraer los intervalos de confianza de 95% (conf.int)
get.conf.int=function(h1) t.test(h1)$conf.int
conf.inth=apply(draws,2,get.conf.int)
conf.inth
#promedio de los intervalos h1
Mh=mean(conf.inth)
dh1=mean(sqrt(var(conf.inth)))
Mh
dh1
sum(conf.inth[,1]<=Mh&conf.inth[,2]>=Mh)
plot(range(conf.inth),c(0,1+n.draw),type="n",          xlab="intervalos          estándar
      bootstrap",ylab="muestra")
for(i in 1:n.draw) lines(conf.inth[,i],rep(i,2),lwd=0.7)
      abline(v=muh, lwd=0.7, lty=0.70)

#Buscando el cubrimiento del intervalo percentil insesgado corregido bootstrap
mum=mean(m1)
mum
desvm1<-sqrt(var(m1))
desvm1
n.draw=100
n=100
draws=matrix(rnorm(n.draw*n,mum,desvm1),n)
draws

#comandos para extraer los intervalos de confianza de 95% (conf.int)
get.conf.int=function(m1) t.test(m1)$conf.int
conf.intm=apply(draws,2,get.conf.int)
conf.intm
#promedio de los intervalos m1
Mm=mean(conf.intm)
dm1=mean(sqrt(var(conf.intm)))
Mm
dm1
sum(conf.intm[,1]<=Mm&conf.intm[,2]>=Mm)
plot(range(conf.intm),c(0,1+n.draw),type="n",  xlab="intervalos  insesgado  corregido
      bootstrap",ylab="muestra")
for(i in 1:n.draw) lines(conf.intm[,i],rep(i,2),lwd=0.7)
      abline(v=mum, lwd=0.7, lty=0.70)

```

#Generando la muestra bootstrap para el índice normal C_{pk} usando la suposición de normalidad.

```
resamples<-lapply(1:1000, function(i) sample(y,replace=T))
r.cpk<-sapply(resamples,cpk)
r.cpk
mean(r.cpk)
desvr<-sqrt(var(r.cpk))
desvr
asim<- skewness(r.cpk)
asim
#ordenando la muestra obtenida
rr.cpk<-sort(r.cpk)
rr.cpk
```

#Generando un intervalo percentil bootstrap.

```
b1<-c(rr.cpk[25],rr.cpk[975])
```

```
b1
```

#Generando un intervalo estándar bootstrap

```
h1<-c(mean(rr.cpk)-1.96*sqrt(var(rr.cpk)),mean(rr.cpk)+1.96*sqrt(var(rr.cpk)))
```

```
h1
```

#Percentil bootstrap sesgo-correctado

```
zo<-qnorm(sum(rr.cpk<cpk(y))/1000)
```

```
P11<-pnorm(2*zo-1.96)
```

```
Pu1<-pnorm(2*zo+1.96)
```

```
a11<-P11*1000
```

```
a21<-Pu1*1000
```

```
a11
```

```
a21
```

```
m1<-c(rr.cpk[a11],rr.cpk[a21])
```

```
m1
```

#Buscando el cubrimiento del intervalo percentil bootstrap

```
mu=mean(b1)
```

```
mu
```

```
desvb1<-sqrt(var(b1))
```

```
desvb1
```

```
n.draw=100
```

```
n=100
```

```
draws=matrix(rnorm(n.draw*n,mu,desvb1),n)
```

#comandos para extraer los intervalos de confianza de 95% (conf.int)

```
get.conf.int=function(b1) t.test(b1)$conf.int
```

```
conf.intb=apply(draws,2,get.conf.int)
```

```
conf.intb
```

#promedio de los intervalos b1

```
Mb=mean(conf.intb)
```

```
db1=mean(sqrt(var(conf.intb)))
```

```
Mb
```

```

db1
sum(conf.intb[1,]<=Mb&conf.intb[2,]>=Mb)
plot(range(conf.intb),c(0,1+n.draw),type="n", xlab="intervalos percentil
  bootstrap",ylab="muestra")
for(i in 1:n.draw) lines(conf.intb[,i],rep(i,2),lwd=0.7)
  abline(v=mu, lwd=0.7, lty=0.70)

#Buscando el cubrimiento del intervalo estándar bootstrap
muh=mean(h1)
muh
desvh1<-sqrt(var(h1))
desvh1
n.draw=100
n=100
draws=matrix(rnorm(n.draw*n,muh,desvh1),n)
draws

#comandos para extraer los intervalos de confianza de 95% (conf.int)
get.conf.int=function(h1) t.test(h1)$conf.int
conf.inth=apply(draws,2,get.conf.int)
conf.inth
#promedio de los intervalos h1
Mh=mean(conf.inth)
dh1=mean(sqrt(var(conf.inth)))
Mh
dh1
sum(conf.inth[1,]<=Mh&conf.inth[2,]>=Mh)
plot(range(conf.inth),c(0,1+n.draw),type="n", xlab="intervalos estándar
  bootstrap",ylab="muestra")
for(i in 1:n.draw) lines(conf.inth[,i],rep(i,2),lwd=0.7)
  abline(v=muh, lwd=0.7, lty=0.70)

#Buscando el cubrimiento del intervalo percentil insesgado corregido bootstrap
mum=mean(m1)
mum
desvm1<-sqrt(var(m1))
desvm1
n.draw=100
n=100
draws=matrix(rnorm(n.draw*n,mum,desvm1),n)
draws

#comandos para extraer los intervalos de confianza de 95% (conf.int)
get.conf.int=function(m1) t.test(m1)$conf.int
conf.intm=apply(draws,2,get.conf.int)
conf.intm
#promedio de los intervalos m1

```

```

Mm=mean(conf.intm)
dm1=mean(sqrt(var(conf.intm)))
Mm
dm1
sum(conf.intm[1,]<=Mm&conf.intm[2,]>=Mm)
plot(range(conf.intm),c(0,1+n.draw),type="n", xlab="intervalos insesgado corregido
  bootstrap",ylab="muestra")
for(i in 1:n.draw) lines(conf.intm[i],rep(i,2),lwd=0.7)
  abline(v=mum, lwd=0.7, lty=0.70)

```

ANEXO 5: Algoritmo genético.

```

% *****Programa para hallar valores en una ecuación dada*****
% *****a través del Algoritmo Genético*****
clc
clear
t = cputime;
Fitnessmayor=[0 0 0 0 0 0 0 0];
Precision=0.0000001 % <- Precision para el ajuste
Max_Iteracion=100 % <- Numero maximo de iteraciones del programa

% *****Primera Generacion(aleaoria)*****
Y(1,:)=dec2bin(999999999,32);
Y(2,:)=dec2bin(999999999,32);
Y(3,:)=dec2bin(99999999,32);
Y(4,:)=dec2bin(9999999,32);
Y(5,:)=dec2bin(999999,32);
Y(6,:)=dec2bin(99999,32);
Y(7,:)=dec2bin(999,32);
Y(8,:)=dec2bin(randi([50000,70000]),32);

% *****
% Y(1,:)=dec2bin(randi([0,70000]),32);
% Y(2,:)=dec2bin(randi([0,70000]),32);
% Y(3,:)=dec2bin(randi([0,70000]),32);
% Y(4,:)=dec2bin(randi([0,70000]),32);
% Y(5,:)=dec2bin(randi([0,70000]),32);
% Y(6,:)=dec2bin(randi([0,70000]),32);
% Y(7,:)=dec2bin(randi([0,70000]),32);
% Y(8,:)=dec2bin(randi([0,10]),32);
Y(1,:)
Y(2,:)
Y(3,:)
Y(4,:)

```

```

Y(5,:)
Y(6,:)
Y(7,:)
Y(8,:)
No_Mutaciones=0;
No_Bit_Mutado=0;

for veces= 1:1:Max_Iteracion
    clear Indices FitOrd Fitness i2 i3 i4 i5 i8 i9 j;
    Tiempo_Minutos=(cputime-t)/60

% *****AJUSTE SELECCIONANDO EL MEJOR VALOR*****
for i2 = 1:1:8

Fitness(i2)=abs((2*(1-
((bin2dec(Y(i2,:)))/1000000))*sqrt(1+2*((bin2dec(Y(i2,:)))/1000000)))...
/(1+3*((bin2dec(Y(i2,:)))/1000000))-0.5645426);
% Fitness(i2)=abs((((bin2dec(Y(i2,:)))/1000000))^2)/((1+ 2*0.5007487)*(1 +
0.5007487)^2) - 0.2338198);
% Fitness(i2)=abs(((bin2dec(Y(i2,:)))/1000000)+ 1.026659/(1+0.5007487))-0.9771234);
% Fitness(i2)=abs((((bin2dec(Y(i2,:)))/1000000))^2) - 4);

end

% *****
% *****Ordenar los Mejores*****
FitOrd=sort(Fitness');
% *****
Fitnessmenor(veces)=(FitOrd(1));
% *Halla la posicion de los cuatro números binarios con mejor Fitness***
Indices=[0 0 0 0 0 0 0];
% clear i3;
for i3 = 1:1:4
    Indices_ref=find(Fitness'==FitOrd(i3));
    Indices(i3)=Indices_ref(1);
end

% *****
% *****Quedarse con los cuatro mejores Fitness*****
for i4 = 1:1:4
    j4=Indices(i4);
    YFit(i4,:)=Y(j4,:);
end

```

```

% *****
% ****Remplazar los cuatro primero por los mejores Fitness****

for i5 = 1:1:4
    Y(i5,:)=YFit(i5,:);
end
% *****
q1a=Y(1,:);
q1b=Y(2,:);
q2a=Y(3,:);
q2b=Y(4,:);
numero_de_bit=10;
% ***** CrossOver 1 con 2*****
i=1;
k=2;
posicion_del_crossover = ceil(numero_de_bit*rand(1));
for j = posicion_del_crossover:1:numero_de_bit
    Y(i,j)= YFit(k,j);
end

% *****
% ***** CrossOver 2 con 1*****
i=2;
k=1;
posicion_del_crossover = ceil(numero_de_bit*rand(1));
for j = posicion_del_crossover:1:numero_de_bit
    Y(i,j)= YFit(k,j);
end

% *****
% ***** CrossOver 3 con 4*****
i=3;
k=4;
posicion_del_crossover = ceil(numero_de_bit*rand(1));
for j = posicion_del_crossover:1:numero_de_bit
    Y(i,j)= YFit(k,j);
end

% *****
% ***** CrossOver 4 con 3*****
i=4;
k=3;
posicion_del_crossover = ceil(numero_de_bit*rand(1));
for j = posicion_del_crossover:1:numero_de_bit

```

```

    Y(i,j)= YFit(k,j);
end

% *****
Y1=Y(1,:);
Y2=Y(2,:);
Y3=Y(3,:);
Y4=Y(4,:);
for i8 = 5:1:8
    Y(i8,:)=YFit(i8-4,:);
end
Y5=Y(5,:);
Y6=Y(6,:);
Y7=Y(7,:);
Y8=Y(8,:);

% ***** GENERACION DEL BIT MUTANTE *****
% bin2dec(Y(i,:))
if((Y(5,:)==Y(6,:))|(Y(5,:)==Y(7,:))|(Y(5,:)==Y(8,:))|(Y(6,:)==Y(7,:))|(Y(6,:)==Y(8,:))|(Y(
7,:)==Y(8,:)))
    No_Mutaciones=No_Mutaciones+1;
    for i9 = 1:1:9
        No_Bit_Mutado=No_Bit_Mutado+1;
        numero_a_mutar = ceil(4*rand(1)); % Número a Mutar los cuatro primeros hijos
        posicion_bit_mutante = ceil((numero_de_bit+10)*rand(1)); % Posición del bit de
Mutación
        bit_mutante = bitget(uint32(bin2dec(Y(numero_a_mutar,:))),posicion_bit_mutante);
        D = bitset(uint32(bin2dec(Y(numero_a_mutar,:))),posicion_bit_mutante,~bit_mutante);
        Y(numero_a_mutar,:)=dec2bin(D,32);
    end
end
if (abs((2*(1 - ((bin2dec(Y(i2,:)))/1000000))*sqrt(1 + 2*((bin2dec(Y(i2,:)))/1000000)))/(1
+ 3*((bin2dec(Y(i2,:)))/1000000))-0.5645426)<=Precision);
% if (abs((((bin2dec(Y(i2,:)))/1000000))^2)/((1+ 2*0.5007487)*(1 + 0.5007487)^2) -
0.2338198)<=Precision);
% if (abs(((bin2dec(Y(i2,:)))/1000000)+ 1.026659/(1+0.5007487)-0.9771234)<=Precision)
% if(abs((((bin2dec(Y(i2,:)))/1000000))^2) - 4)<=Precision)
    break;
end
end
% Fitnessmayor
figure(01)
plot(1:veces,Fitnessmenor,'-r','LineWidth',2.5)

```

```

grid
format long
No_Bit_Mutado
% No_Mutaciones
No_generaciones=veces
respuesta=bin2dec(Y(8,:))/1000000
Tiempo_en_minutos =(cputime-t)/60

```

ANEXO 6: Estimación de ICP mediante la DPG3 con parámetros conocidos y siguiendo los métodos: percentiles de Clements, exacto y percentiles de Burr usando S-Plus.

```

set.seed(38356)
a      <-  0.554
b      <-  1.116
c      <-  0.277
x<-seq(0,2, length=200)
u<-seq(0.005,0.9, len = 200)

# Función de densidad de probabilidad de la DPG3
dpareto <- function(x) (1/b)*( 1 - (a*(x-c)/b))^((1/a)-1)
# Función de distribución acumulada de la DPG3
ppareto <- function(x) (1-( 1 - a*(x-c)/b)^(1/a))
#Función cuantil
qpareto <- function(u) (c + ((b/a)*( 1 - (1 - u)^(a))))
par(mfrow=c(2,2))
plot(x,dpareto(x),type="l")
title("FUNCION DE DISTRIBUCION ACUMULADA", cex.main = 1, font.main= 4,
      col.main= "blue")
plot(x,ppareto(x), type="l")
title("FUNCION DE DENSIDAD DE PROBABILIDAD",cex.main = 1, font.main= 4,
      col.main= "blue")
plot(u,qpareto(u), type="l")
title("Función cuantil", cex.main = 1, font.main= 4, col.main= "blue")
par(mfrow=c(1,1))
rpareto <- function(n) qpareto(runif(n))
y      <-  rpareto(100)
y
hist(y)
qpareto(0.95)
qpareto(c(0.99865,0.5,0.135))
quantile(y,c(0.99865,0.5,0.135))
qpareto(0.99865)
qpareto(0.5)
qpareto(0.135)
summary(y)
min(y)
max(y)

```

```

#Quantiles
Lp<-quantile(y,0.00135)
Med<-quantile(y,0.5)
Up<-quantile(y,0.99865)
Lp
Med
Up
medidas<-(c(mean(y),sqrt(var(y)),skewness(y),kurtosis(y)))
Lt=0.43
Medidas

```

```

#Metodo de Clements
#Para generar el Ut
Cpu=0.5
Ut<-Cpu*(Up-Med)+Med
Ut
Cp=(Ut-Lt)/(Up-Lp)
Cpks=(Ut-Med)/(Up -Med)
Cpki=(Med-Lt)/(Med-Lp)
Cpks
Cpki
Cpk=min(Cpks,Cpki)
Cp
Cpk

```

```

#Metodo exacto
a      <-  0.554
b      <-  1.116
c      <-  0.277
#Para generar el Ut
Cpu=2
Ut<-Cpu*(Up-Med)+Med
Ut
Lp0135= (c+((b/a)*(1-(1-0.00135)^(a))))
M05= (c+((b/a)*(1-(1-0.5)^(a))))
Up99865= (c+((b/a)*(1-(1-0.99865)^(a))))
Lp0135
M05
Up99865
Cp=(Ut-Lt)/(Up99865-Lp0135)
Cpks=(Ut-M05)/(Up99865-M05)
Cpki=(M05-Lt)/(M05-Lp0135)
Cp
Cpks
Cpki

```

www.bdigital.ula.ve

```

#Método de Percentiles de Burr:
Lp<-quantile(y,0.00135)
Med<-quantile(y,0.5)
Up<-quantile(y,0.99865)
sk<-skewness(y)
k<-kurtosis(y)
Lt=0.43
medidas
#momentos estandarizados de asimetria y kurtosis
alp3<-(n-2)*sk/sqrt(n*(n-1))
alp4<-((n-2)*(n-3)*k/(n^2-1))+(3*(n-1)/(n+1))
alp3
alp4
# VALORES TABULADOS
#Para construir los LP(con BZ0.0135) M(con BZ0.5) Up(con BZ0.99865)buscados en la tabla
BZ0135=-0.650
BZ5=-0.516
BZ99865=4.877

#Para construir los LP(con Z0.0135) M(con Z0.5) Up(con Z0.99865)buscados en la tabla
CZ0135=-0.758
CZ5=-0.416
CZ99865=5.012
www.bdigital.ula.ve
#ESTIMACIÓN DE LOS PERCENTILES ESTANDARIZADOS PARA BURR Y PARA
  CLEMENTS

#procedimiento para los percentiles de Burr
BLp<-function(y)(mean(y)+ sqrt(var(y))*BZ0135)
BM<-function(y)(mean(y)+ sqrt(var(y))*BZ5)
BUp<-function(y)(mean(y)+ sqrt(var(y))*BZ99865)
BLp(y)
BM(y)
BUp(y)

#procedimiento para los percentiles de Clements
CLp<-function(y)(mean(y)+ sqrt(var(y))*CZ0135)
CM<-function(y)(mean(y)+ sqrt(var(y))*CZ5)
CUp<-function(y)(mean(y)+ sqrt(var(y))*CZ99865)
CLp(y)
CM(y)
CUp(y)

#ESTIMACIÓN DE LOS ÍNDICES NO NORMALES Clement y Burr) USANDO LAS
  FÓRMULAS RESPECTIVAS.
Lt=0,43

```

#ICP de Burr

```
BCp<-function(y) (Ut-Lt)/(BUp(y)-BLp(y))
BCpu<-function(y)(Ut-BM(y))/(BUp(y)-BM(y))
BCpl<-function(y) (BM(y)-Lt)/(BM(y)-BLp(y))
BCpk<-function(y) min(BCpl(y),BCpu(y))
```

#ICP de Clements metodologia Liu-Chen

```
CCp<-function(y) (Ut-Lt)/(CUp(y)-CLp(y))
CCpu<-function(y)(Ut-CM(y))/(CUp(y)-CM(y))
CCpl<-function(y)(CM(y)-Lt)/(CM(y)-CLp(y))
CCpk<-function(y) min(CCpu(y),CCpl(y))
```

#Resultados

```
BCp(y)
BCpk(y)
CCp(y)
CCpk(y)
```

#Indice Cpk con distribución normal

#Para generar el Ut

Cpu=2

Ut<-Cpu*(Up-Med)+Med

Ut

```
cp<-function(y) (Ut-Lt)/(6*sqrt(var(y)))
cpl<-function(y)(mean(y)-Lt)/(3*sqrt(var(y)))
cpu<-function(y)(Ut-mean(y))/(3*sqrt(var(y)))
cpk<-function(y) min(cpu(y),cpl(y))
cp(y)
cpl(y)
cpu(y)
cpk(y)
```