



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Máster Universitario en Ingeniería de Análisis de Datos,
Mejora de Procesos y Toma de Decisiones

Trabajo Final de Máster

**Diseño óptimo de gráficos de control no
paramétricos para la dispersión basados
en el estadístico de signo en presencia de
errores de redondeo**

Autor: Javier Porcel Marí

Tutores: Vicent Giner Bosch

Philippe Castagliola

Curso: 2020-2021

Agradecimientos

Quiero dar las gracias a mi tutor D. Vicent Giner Bosch por haberme acompañado y aconsejado a lo largo de la realización del presente trabajo, y a mi otro tutor D. Philippe Castagliola por su colaboración.

Dedicatoria

A mi padre y a mi madre, Félix y María Dolores; y a mi hermano pequeño, Marcos.

Resumen

Hoy en día, los gráficos de control son una herramienta potente y útil para alcanzar estándares de calidad elevada en una industria altamente competitiva, pues nos permiten controlar y mejorar los procesos industriales.

De este gran interés surge la motivación de este trabajo final de máster, que consiste en optimizar, desde un punto de vista estadístico y computacional, un gráfico de control Shewhart no paramétrico basado en el estadístico de signo para detectar cambios en la dispersión de una característica de calidad continua. En concreto, se aborda el estudio del caso en el que se producen errores debidos a la resolución del instrumento de medida.

En primer lugar, se define el estadístico de signo para la dispersión y se formaliza matemáticamente la presencia de errores de redondeo asociados al instrumento de medida. A partir de aquí, se desarrollan dos algoritmos enumerativos para la determinación óptima de los parámetros del gráfico objeto de estudio: uno para el caso en el que se asume que no hay errores de redondeo, y otro teniendo en cuenta la presencia de este tipo de errores.

Posteriormente, se realizan experiencias numéricas para analizar, en primer lugar, el comportamiento del gráfico optimizado para escenarios sin errores de redondeo cuando realmente sí se producen este tipo de errores; como cabía prever, se constata una degradación del desempeño esperado del gráfico. En segundo lugar, se evalúan los resultados de *reoptimizar* el gráfico para un escenario con errores de redondeo; se observa que es posible obtener gráficos que tengan en cuenta los errores de redondeo manteniendo niveles de rendimiento similares a los esperados.

Por último, las experiencias numéricas llevadas a cabo revelan la importancia de realizar un estudio de las características de la distribución de probabilidad subyacente (en particular, de su curtosis) para obtener estimaciones ajustadas del desempeño del gráfico.

Palabras clave: gráficos de control no paramétricos; dispersión; estadístico de signo; error de redondeo; resolución; distribución de Johnson; optimización en ingeniería; optimización en control de la calidad.

Resum

Hui dia, els gràfics de control són una eina potent i útil per a aconseguir estàndards de qualitat elevada en una indústria altament competitiva, ja que ens permeten controlar i millorar els processos industrials.

D'aquest gran interès sorgeix la motivació d'aquest treball final de màster, que consisteix a optimitzar, des d'un punt de vista estadístic i computacional, un gràfic de control Shewhart no paramètric basat en l'estadístic de signe per a detectar canvis en la dispersió d'una característica de qualitat contínua. En concret, s'aborda l'estudi del cas en el qual es produeixen errors deguts a la resolució de l'instrument de mesura.

En primer lloc, es defineix l'estadístic de signe per a la dispersió i es formalitza matemàticament la presència d'errors d'arrodoniment associats a l'instrument de mesura. A partir d'ací, es desenvolupen dos algorismes enumeratius per a la determinació òptima dels paràmetres del gràfic objecte d'estudi: un per al cas en el qual s'assumeix que no hi ha errors d'arrodoniment, i un altre tenint en compte la presència d'aquests tipus d'errors.

Posteriorment, es realitzen experiències numèriques per a analitzar, en primer lloc, el comportament del gràfic optimitzat per a escenaris sense errors d'arrodoniment quan realment sí que es produeixen aquest tipus d'errors; com cabia preveure, es constata una degradació del rendiment esperat del gràfic. En segon lloc, s'avaluen els resultats de *reoptimitzar* el gràfic per a un escenari amb errors d'arrodoniment; s'observa que és possible obtindre gràfics que tinguen en compte els errors d'arrodoniment tot mantenint nivells de rendiment similars als esperats.

Finalment, les experiències numèriques dutes a terme revelen la importància de realitzar un estudi de les característiques de la distribució de probabilitat subjacent (en particular, de la seua curtosi) per a obtindre estimacions ajustades del rendiment del gràfic.

Paraules clau: gràfics de control no paramètrics; dispersió; estadístic de signe; error d'arrodoniment; resolució; distribució de Johnson; optimització en enginyeria; optimització en control de la qualitat.

Abstract

Nowadays, control charts are a powerful and useful tool to achieve high quality standards in a highly competitive industry, since they allow us to monitor and improve industrial processes.

From this great interest arises the motivation of this master's thesis, which consists of optimising, from a statistical and computational point of view, a non-parametric Shewhart control chart based on the sign statistic aiming at detecting changes in the dispersion of a continuous quality characteristic. In particular, the study of the case in which errors occur due to the resolution of the measuring instrument is addressed.

First, the sign statistic for dispersion is defined and the presence of rounding errors associated with the measuring instrument is mathematically formalised. From here, two enumerative algorithms are developed for the optimal determination of the parameters of the graph under study: one for the case in which it is assumed that there are no rounding errors, and another one taking into account the presence of this type of errors.

Afterwards, numerical experiments are carried out to analyse, firstly, the behaviour of the graph optimised for scenarios without rounding errors when rounding errors do occur; as expected, a degradation of the performance of the graph is observed. Secondly, the results of *re-optimising* the graph for a scenario with rounding errors are evaluated; it is observed that it is possible to obtain graphs that take into account rounding errors while maintaining performance levels similar to those expected.

Finally, the numerical experiments carried out reveal the importance of performing a study of the characteristics of the underlying probability distribution (in particular, its kurtosis) in order to obtain adjusted estimates of the chart's performance.

Keywords: nonparametric control charts; dispersion; sign statistic; rounding-off error; resolution; Johnson distribution; optimisation in engineering; optimisation in quality control.

Contenido

1.	Introducción	1
1.1.	Control estadístico de la calidad	1
1.2.	Gráficos de control.....	2
1.2.1.	Medidas de desempeño de un gráfico de control.....	4
1.2.2.	Parámetros de diseño de un gráfico de control	5
1.2.3.	Gráficos de control para la dispersión	5
1.2.4.	Gráficos de control no paramétricos.....	6
1.2.5.	Errores en la medición	7
1.3.	Antecedentes y motivación	7
1.4.	Objetivos	8
1.5.	Estructura del trabajo.....	8
2.	Marco teórico.....	9
2.1.	Estadístico de signo para la dispersión	9
2.2.	Gráfico de control no paramétrico de Pawar et al. (2018).....	11
2.3.	Errores de redondeo en el instrumento de medida	12
2.4.	Distribuciones de Johnson	13
2.4.1.	Definición.....	13
2.4.2.	Selección de distribuciones para un banco de pruebas.....	15
2.4.3.	Determinación de los parámetros de una distribución de Johnson con un cambio de dispersión τ	17
3.	Metodología.....	18
3.1.	Notación	18
3.2.	Etapa 1: diseño óptimo del gráfico de control para la dispersión para mediciones sin errores de redondeo.....	19
3.3.	Etapa 2: cálculo del ARL_0 y ARL_1 real ante la presencia de <i>empates</i> en los gráficos diseñados.....	24
3.4.	Etapa 3: diseño óptimo del gráfico de control para la dispersión en presencia de empates	26
4.	Resultados.....	29
4.1.	Análisis de los resultados obtenidos para la situación 1	29
4.2.	Análisis de los resultados obtenidos para la situación 2	31
4.3.	Análisis de los resultados obtenidos para la situación 3	33
4.4.	Análisis de la situación 2 frente a la situación 1	35
4.5.	Análisis de la situación 3 frente a la situación 2.....	38
4.6.	Análisis comparativo conjunto de las tres situaciones	41

4.7. Análisis de la situación 3 frente a la situación 1	42
5. Conclusiones	45
Bibliografía.....	47
Anexos	50
Anexo A. Código de Scilab.....	50
Anexo B. Archivo Excel con los resultados numéricos	51
Anexo C. Tablas de los ANOVAS de las tres situaciones y de la comparación de la situación 3 frente a la situación 1	52
Anexo D. Tablas de medias para cada una de las tres situaciones y para la comparación de la situación 3 frente a la situación 1.....	54

1.Introducción

A lo largo de la historia, el ser humano se ha preocupado por la calidad de los bienes que produce, fabrica y consume. Esto se debe a que los productos y servicios de calidad permiten que el cliente o consumidor quede satisfecho. En particular, mediante una buena adecuación al uso y el cumplimiento de determinadas especificaciones o requisitos, el producto puede cubrir las necesidades y expectativas del cliente.

Montgomery (2012) llega a enumerar hasta ocho dimensiones de la calidad: desempeño (¿el producto cumple su función?), fiabilidad (¿con qué frecuencia falla el producto?), durabilidad (¿cuánto dura el producto?), mantenibilidad (¿el producto es fácilmente reparable?), estética (¿el producto es bonito?), características (¿qué hace el producto?), calidad percibida (¿el producto tiene buena reputación?) y conformidad con las especificaciones (¿el producto se ha fabricado acorde con los estándares del diseño?).

Como parte de la gestión de la calidad en una empresa, el control estadístico de la calidad proporciona los instrumentos adecuados para medir y mejorar la calidad de los procesos productivos (Alfaro-Cortés et al., 2004). En concreto, el control estadístico de procesos y, en particular, su herramienta estrella, los gráficos de control, permiten monitorizar el comportamiento de un proceso para determinar si opera de acuerdo con lo esperado. Este aspecto es fundamental desde el punto de vista de la calidad, en el sentido de que no puede asegurarse la calidad de un proceso fuera de control, esto es, un proceso cuyo comportamiento no se corresponde con el esperado.

En las próximas secciones de este capítulo introductorio se desarrollan estos y otros conceptos, para, posteriormente, motivar y definir el objetivo del presente trabajo.

1.1. Control estadístico de la calidad

Según Juran y Godfrey (2001), la ingeniería de calidad se define como el conjunto de actividades de una empresa u organización llevadas a cabo para asegurar que las características de la calidad (esto es, las propiedades de un producto que tienen que ver con su calidad) alcancen los niveles requeridos, y se divide en tres partes: (1) la planificación de la calidad, (2) el control de la calidad y (3) la mejora de la calidad.

La *planificación de la calidad* consiste en especificar qué estándares debe cumplir el producto a fabricar, y cuál debe ser el diseño al que se debe de ajustar el mismo.

El *control estadístico de la calidad* trata de verificar mediante técnicas estadísticas adecuadas que el producto en cuestión se está fabricando acorde con el diseño y con los requerimientos fijados en la fase de planificación de la calidad.

Por último, la *mejora de la calidad* aborda qué medidas tomar para incrementar la calidad del producto, consiguiendo así una mejora continua.

Dentro del control estadístico de la calidad, ámbito en el cual se encuadra el presente trabajo, pueden distinguirse tres tareas, que se realizan de manera cíclica y continua: la *medición* de los parámetros de calidad, la *comparación* de los valores medidos con la norma, y la *decisión* de las acciones correctoras necesarias.

En particular, el control estadístico de procesos es el área del control estadístico de la calidad que proporciona las herramientas para asegurar que el proceso opera de manera estable y con poca variabilidad alrededor de los valores esperados para las características de calidad que lo componen (Montgomery, 2012). Dichas herramientas pueden ser tan sencillas como los histogramas, el diagrama de tallo y hojas o los diagramas de causa-efecto; y al mismo tiempo tan potentes como los gráficos de control (véase la sección 1.2).

Además de las técnicas mencionadas, existen otras muchas más modernas y complejas que pueden ser de gran utilidad para el control estadístico de procesos, tales como el control estadístico de procesos multivariantes basado en PCA (Ferrer-Riquelme, 2009) o el control estadístico de procesos basado en redes neuronales (Zorriassatine y Tannock, 1998).

1.2. Gráficos de control

Los gráficos de control (propuestos por primera vez por W.A. Shewhart entre la tercera y cuarta década del siglo XX) son una potente herramienta para el control estadístico de procesos. En esencia, consisten en la monitorización de un estadístico muestral a lo largo del tiempo, realizándose varias mediciones y observando si estas mediciones se encuentran dentro de unos rangos tolerables llamados límites de control. En otras palabras, tal y como el propio Shewhart los definió, se trata de “un proceso dinámico, continuo y autocorrectivo diseñado con el propósito de conseguir y mantener un estado de control estadístico” (Shewhart, 1986).

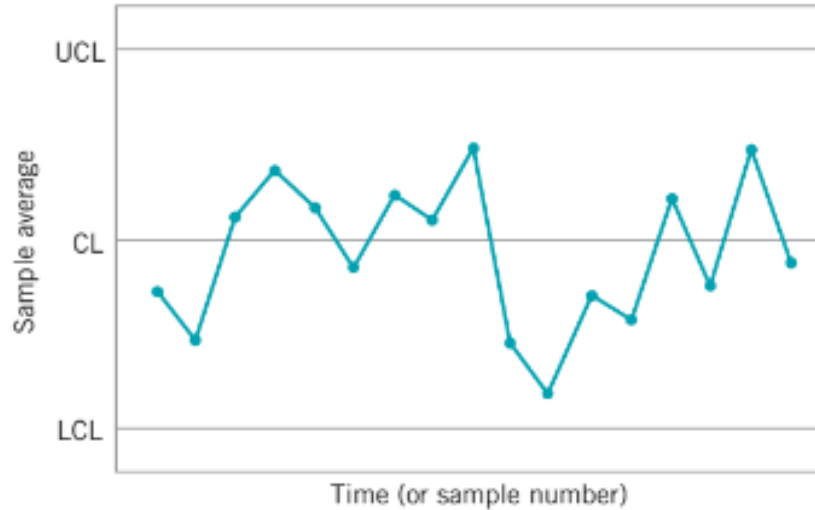


Figura 1.1. Ejemplo de gráfico de control. Fuente: Montgomery (2012).

La figura 1.1 muestra un gráfico de control sencillo. En él, el estadístico objeto de monitorización es la media muestral, cuyo valor objetivo o esperado se representa mediante la *línea central* del gráfico (o CL, por sus siglas en inglés). El gráfico se completa con los *límites de control inferior y superior* (LCL y UCL, respectivamente, por sus siglas en inglés). Cada punto del gráfico corresponde al resultado de calcular y graficar el valor del estadístico monitorizado (en este caso, la media muestral \bar{X}) a partir de una muestra aleatoria de n mediciones de la característica de calidad objeto de estudio. Cada muestra se supone independiente de las anteriores.

Cuando se obtiene un punto fuera de los límites de control, se dice que se ha generado una *señal de falta de control*. Lo deseable al diseñar y utilizar un gráfico de control es que, cuando el proceso opera *bajo control* estadístico (es decir, cuando la característica de calidad de interés presenta una distribución constante a lo largo del tiempo o, lo que es lo mismo, su variabilidad es debida únicamente a *causas fortuitas*), la probabilidad de obtener una señal de falta de control (lo que es conocido como la *tasa de falsa alarma*) sea muy pequeña; y al contrario, esta probabilidad debe ser alta cuando el proceso está *fuera de control* (esto es, cuando su comportamiento se ve afectado por causas de variabilidad no fortuitas o *especiales*, también llamadas *asignables*). En cierto modo, un gráfico de control funciona como un contraste de hipótesis que se ejecuta cada vez que se extrae y procesa una nueva muestra. Desarrollamos estos conceptos en la sección siguiente.

1.2.1. Medidas de desempeño de un gráfico de control

Tal como se ha mencionado, en el diseño de un gráfico de control es importante que la cantidad de puntos que se salen de los límites de control sea pequeña cuando el proceso se considera bajo control; es decir, que se produzcan un número reducido de falsas alarmas. Para medir la frecuencia con la que se dan estas falsas alarmas, puede utilizarse la *longitud promedio de racha* bajo control o ARL_0 , que nos indica, en promedio, cada cuántos puntos o muestras se obtiene un punto fuera de los límites de control, asumiendo que el proceso se encuentra bajo control estadístico. Matemáticamente, en el caso de un gráfico clásico o tipo Shewhart (entendemos que un gráfico es de este tipo cuando cada muestra es independiente de las anteriores y presenta límites de control y tamaño muestral constantes a lo largo del tiempo), esta medida cumple la ecuación (1.1):

$$ARL_0 = \frac{1}{\alpha} \quad (1.1)$$

donde α corresponde a la probabilidad de un error de primera especie o de falsa alarma, es decir, que un proceso que está bajo control se comporte como si no lo estuviera.

Por otro lado, no solo es importante que el número de puntos que queden fuera de los límites de control no sean numerosos cuando el proceso está bajo control (de hecho, lo ideal sería $ARL_0 \rightarrow \infty$ cuando el proceso está bajo control) sino que, al mismo tiempo, es deseable que cuando el proceso se encuentra fuera de control no tarde en que algún punto sobresalga de los límites de control para poder así 'alarmarnos' y revisar el proceso para volver a tomar el control de este. Con este fin, se define el parámetro ARL_1 que nos indica cuantos puntos habrá, en promedio, en el gráfico hasta que uno de ellos esté fuera de los límites de control desde que el proceso se salió de control. En un gráfico clásico o tipo Shewhart, el ARL_1 cumple la ecuación (1.2):

$$ARL_1 = \frac{1}{p} = \frac{1}{1 - \beta} \quad (1.2)$$

donde p corresponde a la potencia del gráfico, definida como la probabilidad de que si el proceso está fuera de control se detecte como tal en un momento dado, mientras que β es la probabilidad de error de segunda especie, es decir, β es la probabilidad de que el proceso se comporte como bajo control cuando no lo está.

En definitiva, el diseño de un gráfico de control se realiza teniendo presente que su comportamiento sea el deseado, habitualmente en términos de ARL_0 y ARL_1 (o equivalentemente, de α y β), buscando, por tanto, un equilibrio entre especificidad y potencia (o sensibilidad).

1.2.2. Parámetros de diseño de un gráfico de control

Cuando se habla de diseñar un gráfico de control de forma que su desempeño sea el deseado, nos referimos a definir convenientemente los llamados *parámetros de diseño* del gráfico. Los parámetros de diseño son aquellos valores numéricos involucrados en la definición del gráfico, que son configurables por el usuario final y que influyen en el valor de las medidas de desempeño del gráfico comentadas arriba (esto es, que determinan el comportamiento del gráfico).

En el caso de un gráfico de Shewhart tradicional como el mostrado en la figura 1.1, los parámetros de diseño serían los límites de control *LCL* y *UCL*.

Hay que tener en cuenta que, según la situación objeto de monitorización (y de manera análoga a lo que sucede con los contrastes de hipótesis), algunos gráficos se definen como *unilaterales*; es decir, con solo un límite de control (el superior o el inferior).

1.2.3. Gráficos de control para la dispersión

El gráfico de control más habitual y sencillo es el que tiene como objeto monitorizar la media muestral (véase la figura 1.1). Está pensado para detectar desviaciones en la *posición* o *localización* del proceso.

En ocasiones, en cambio, el interés está en monitorizar la *dispersión* o la *variabilidad* de la característica de calidad objeto de interés (en lugar de o además de su posición). De hecho, Collani y Sheil (1989) (citados por Huang y Chen, 2005) destacan cómo algunas situaciones (tales como materias primas defectuosas, operarios no cualificados y degradación de la maquinaria) pueden conducir a un aumento no deseado de la variabilidad del proceso sin influir necesariamente en la media de la característica de calidad. Además, tal y como mencionan Lowry et al. (1995) y Nelson (1990) entre otros, una disminución de la dispersión puede indicar formas de mejorar el proceso. Por todos estos motivos, puede resultar de interés realizar un seguimiento a lo largo del tiempo de la dispersión del proceso.

En este trabajo se busca diseñar de forma óptima un gráfico de control para la dispersión. Para contextualizar esta tarea históricamente, es importante mencionar que los primeros gráficos de control para la dispersión fueron elaborados y presentados por Shewhart (1931), quien diseñó el gráfico R (basado en monitorizar el rango) y el gráfico S (basado en monitorizar la desviación típica muestral). Posteriormente, se diseñó la versión de *sumas acumuladas* (*cumulative sum*; CUSUM) del gráfico S por Page (1963) y por Hawkins (1981); y también la versión de *medias móviles ponderadas exponencialmente* (*exponentially weighted moving average*; EWMA) del gráfico S por Crowder y Hamilton (1992). Más

recientemente, He y Grigoryan (2002) diseñaron gráficos de control de doble muestreo para la desviación típica y Castagliola et al. (2007) diseñaron un gráfico EWMA adaptativo para la dispersión. Otros trabajos más recientes que merece la pena mencionar son el de Riaz (2008), quien diseñó un gráfico basado en monitorizar el rango intercuartílico; los gráficos para el coeficiente de variación de Kang et al. (2007) y Yeong et al. (2017), entre otros; y el artículo de Pawar et al. (2018), quienes diseñan el primer gráfico para la dispersión basado en el estadístico de signo.

1.2.4. Gráficos de control no paramétricos

Los gráficos de control paramétricos son aquellos gráficos que presuponen que la característica de calidad que se desea monitorizar posee una distribución probabilística subyacente conocida; habitualmente, una distribución normal o gaussiana.

Sin embargo, en ocasiones, la hipótesis de normalidad no se cumple, o se desconoce la distribución concreta que sigue el proceso, por lo que se vuelve necesario diseñar gráficos de control no paramétricos que puedan lidiar con esta problemática, ya que de no hacerlo se puede llegar a conclusiones erróneas. Fruto de este interés, existen esfuerzos académicos e investigadores en esta temática, tales como los trabajos de Langenberg y Iglewicz (1986), Alloway y Raghavachari (1991) y Yourstone y Zimmer (1992).

Según Chakraborti et al. (2001), un gráfico de control es no paramétrico “si la distribución del *ARL* bajo control es la misma para cualquier distribución continua”. Es decir, un gráfico de control se considera formalmente no paramétrico si sus propiedades cuando el proceso está bajo control no dependen de la distribución subyacente. Si bien es cierto que los métodos no paramétricos están en desventaja ante los paramétricos cuando los datos estudiados se comportan normal (tanto en control de calidad como en inferencia estadística en general), presentan también la ventaja de ser más robustos en presencia de observaciones anómalas y pueden ofrecer un comportamiento más eficiente que las técnicas paramétricas cuando se trabaja con distribuciones con elevada asimetría (Gibbons y Chakraborti, 2003).

Entre los numerosos gráficos de control no paramétricos existentes en la literatura, destacamos uno propuesto por Pawar et al. (2018). Se trata de un gráfico no paramétrico para la dispersión, basado en una redefinición del estadístico de signo (el cual, habitualmente, se relaciona con la mediana de una distribución con el fin de monitorizar la localización o posición de un proceso, y no su dispersión).

En las secciones 2.1 y 2.2 se describirán esta definición del estadístico de signo para la dispersión y también el gráfico de Pawar et al., que será el punto de partida de nuestro trabajo.

1.2.5. Errores en la medición

En las últimas décadas, existe bastante interés en la literatura en control estadístico de procesos por el estudio del impacto que tienen distintos tipos de error en el diseño y el rendimiento de los gráficos de control (para más detalles, véase el trabajo de Maleki et al., 2017).

Las diferentes fuentes de error que pueden darse al realizar una medición son detalladas, entre otros, por Linna y Woodall (2001). Según estos autores, son las siguientes:

- Error lineal de *sesgo* (*bias-linearity error*): error causado por el calibrado del instrumento de medida. Indica cuál es la diferencia, en promedio, entre el valor real y el valor resultante de la medición.
- Error de *precisión*: error relacionado con la variabilidad del instrumento de medida al realizar varias veces la misma medición.
- Error de *redondeo*: error relacionado con la *resolución* del instrumento de medida.

El presente trabajo se centra en analizar el efecto del error de redondeo causado por la resolución del instrumento de medida sobre el diseño de un determinado gráfico de control; los otros dos tipos de error se considerarán nulos. El tratamiento matemático de este tipo de error se introducirá en la sección 2.3.

1.3. Antecedentes y motivación

Nuestro trabajo consiste en el diseño óptimo de un gráfico de control de Shewhart no paramétrico para la dispersión basado en el estadístico de signo y teniendo en cuenta la presencia de errores de redondeo. Se partirá del trabajo realizado por Pawar et al. (2018), quienes diseñaron por primera vez un gráfico de control para la dispersión basado en el estadístico de signo. Se busca dar respuesta a las siguientes limitaciones que posee el gráfico de control introducido por Pawar et al.:

- Solo permite detectar incrementos en la dispersión, mientras que en este trabajo se busca poder detectar tanto incrementos como decrementos.
- No proporciona una metodología para determinar de manera óptima los parámetros del gráfico.
- No tiene en cuenta los posibles errores de redondeo.

Algunos trabajos previos, como el de Castagliola et al. (2020), evidencian la necesidad de tener en cuenta este tipo de error, si bien su trabajo se centra en estudiar el efecto de la resolución del instrumento de medida sobre un gráfico no paramétrico basado también en el estadístico de signo, no para la dispersión (como es de interés en nuestro caso) sino para la posición.

Esto justifica la actualidad y la utilidad de la temática abordada en este trabajo final de máster.

1.4. Objetivos

Este trabajo final de máster pretende alcanzar los siguientes objetivos:

1. Estudiar el gráfico de control no paramétrico para la dispersión basado en el estadístico de signo de Pawar et al. (2018) e implementar un procedimiento para la determinación óptima de sus parámetros.
2. Analizar el impacto de la existencia de errores de redondeo en el instrumento de medida sobre el comportamiento esperado del gráfico no paramétrico para la dispersión objeto de estudio.
3. Diseñar un procedimiento para la determinación óptima de los parámetros del gráfico objeto de estudio asumiendo la presencia de errores de redondeo, y estudiar el comportamiento del gráfico resultante frente a lo obtenido al ignorar los errores de redondeo.

1.5. Estructura del trabajo

A partir de aquí, el trabajo consta de cuatro capítulos más, sin contar los anexos.

En el segundo capítulo, se introducen algunos conceptos teóricos y conocimiento previo sobre los que se basa nuestra propuesta.

En el tercer capítulo se detalla la metodología que se ha seguido en el trabajo. De esta forma, se describen los algoritmos creados para optimizar el gráfico objeto de estudio sin y con presencia de redondeos, y también, los fundamentos matemáticos que se esconden detrás de estos.

En el cuarto capítulo se muestran los resultados de aplicar nuestra propuesta sobre un conjunto de condiciones iniciales representativas de las diferentes situaciones que pueden darse en la práctica. En concreto, se evalúa el comportamiento del gráfico optimizado sin tener en cuenta los errores de redondeo en presencia de éstos; y los resultados de la *reoptimización*.

En el quinto y último capítulo se muestran las conclusiones que pueden extraerse del trabajo realizado.

2. Marco teórico

Antes de explicar la metodología que se ha seguido en este trabajo, es necesario introducir una serie de conceptos: el estadístico de signo (estadístico utilizado en el gráfico de control que se busca optimizar), el gráfico de Pawar et al. (2018) que sirve de punto de partida para este trabajo; una formalización matemática para los errores de redondeo; y la familia de distribuciones de Johnson, las cuales se utilizan como distribución de probabilidad para la característica de calidad que se pretende monitorizar en la experiencia computacional posterior para evaluar el comportamiento del gráfico propuesto.

2.1. Estadístico de signo para la dispersión

El estadístico de signo para la dispersión es aquel sobre el que se basará la construcción del gráfico de control que se estudiará en este trabajo. Este estadístico fue descrito por primera vez por Bradley (1968) y posteriormente revisado por Amin et al. (1995). Tras realizar n mediciones $\{X_1, \dots, X_n\}$ de la variable aleatoria X , se define U_j (donde $j = 1, \dots, n$) mediante la ecuación (2.1).

$$U_j = \begin{cases} 1, & \text{si } X_j < I_L \text{ o } X_j > I_U \\ 0, & \text{si } X_j = I_L \text{ o } X_j = I_U \\ -1, & \text{si } I_L < X_j < I_U \end{cases} \quad (2.1)$$

Los umbrales o límites I_L e I_U se definen de forma que I_L deja a su izquierda una probabilidad de $p_0/2$ y I_U deja a su derecha una probabilidad también de $p_0/2$. Es decir:

$$\Pr(X \leq I_L) = p_0/2 \quad (2.2)$$

$$\Pr(X > I_U) = p_0/2 \quad (2.3)$$

donde $0 < p_0 < 1$ y $I_L < I_U$. De este modo, p_0 corresponde a la probabilidad de que la medición sobre la variable aleatoria no se encuentre dentro del intervalo $[I_L, I_U]$.

Siempre y cuando X sea continua, el valor $U_j = 0$ será imposible, pues siempre habrá algún decimal que consiga diferenciar entre X_j y I_L o X_j y I_U .

De esta forma, el estadístico de signo para la dispersión se define como la ecuación (2.4).

$$U_t = \sum_{j=1}^n U_j \quad (2.4)$$

De las ecuaciones mostradas, es fácil deducir que el estadístico de signo U_t puede tomar $n + 1$ valores cuando X es una variable continua, concretamente puede valer $\{-n, -n + 2, \dots, 0, \dots, n - 2, n\}$; mientras que, si X es una variable discreta que da lugar a situaciones en que $U_j = 0$, entonces U_t puede tomar $2n + 1$ valores, concretamente, $\{-n, -n + 1, \dots, 0, \dots, n - 1, n\}$.

Asumiendo que X sea una variable aleatoria continua, la distribución que se relaciona con este estadístico es una distribución binomial. Esta asociación se produce fruto de la ecuación (2.5).

$$V = \frac{U + n}{2} \sim Bi(n, p_0) \quad (2.5)$$

Teniendo en cuenta esta relación, se obtiene el valor esperado del estadístico U (ecuación (2.6)) y la varianza de este (ecuación (2.7)):

$$E(U) = n(2p_0 - 1) \quad (2.6)$$

$$\sqrt{Var(U)} = 2\sqrt{np_0(1 - p_0)} \quad (2.7)$$

Habitualmente se utiliza $p_0 = 0.5$; en este caso se tiene que I_L y I_U coinciden con el primer y el tercer cuartil de X , respectivamente (esta es, concretamente, la propuesta de Amin et al, 1995). Cabe notar que es únicamente en esta situación concreta en la que se da que $E(U) = 0$.

En la medida en que la cantidad de observaciones que se transforman en 1 o en -1 según la ecuación (2.1) depende de su posición en relación con el intervalo $[I_L, I_U]$, el estadístico de signo que acabamos de describir puede ser visto como una medida de la dispersión de una variable aleatoria y puede ser usado para monitorizar dicha dispersión a lo largo del tiempo.

2.2. Gráfico de control no paramétrico de Pawar et al. (2018)

Pawar et al. (2018), basándose en un trabajo previo de Amin et al. (1995), diseñaron un gráfico de control no paramétrico para la dispersión basado en el estadístico de signo que acabamos de describir, el cual no tiene en cuenta posibles errores de redondeo ni de otro tipo (véase la subsección 1.2.3). En esta subsección describimos dicho gráfico, ya que será la base de nuestra propuesta.

El gráfico de control de Pawar et al. monitoriza una característica de calidad continua X que se obtiene al realizar muestreos aleatorios de tamaño constante n equiespaciados en el tiempo. Para cada instante t , se calcula el estadístico de signo U_t según la ecuación (2.1) con el fin de realizar un seguimiento de la dispersión de esta variable aleatoria a lo largo del tiempo.

En la ecuación (2.1), los límites I_L e I_U se definen de acuerdo con las ecuaciones (2.2) y (2.3), asumiendo que X se encuentra bajo control. Es decir, se considerará que p_0 corresponde a la probabilidad de que una observación de la característica de calidad X esté fuera del intervalo $[I_L, I_U]$ cuando la dispersión de esta coincide con la esperada o deseada. Esto es:

$$\Pr(I_L \leq X \leq I_U \mid \sigma = \sigma_0) = 1 - p_0 \quad (2.8)$$

Siendo σ_0 la desviación típica de X cuando se encuentra bajo control.

Como ya se ha mencionado en la sección 1.3, Pawar et al. (2018) solo se centraron en detectar incrementos de la dispersión. Por tanto, solo utilizan el límite de control superior o el UCL , de sus siglas en inglés: *Upper Control Limit*. Si se buscará determinar un decremento en la desviación, que no es el caso de Pawar et al., sería necesario determinar un límite de control inferior o LCL , de sus siglas en inglés: *Lower Control Limit*.

Cabe destacar, además, que Pawar et al. no proporcionan ningún procedimiento para determinar de manera óptima el valor de los parámetros del gráfico p_0 y UCL , aunque sí que muestran resultados del desempeño del gráfico para diferentes valores de estos parámetros.

2.3. Errores de redondeo en el instrumento de medida

En este trabajo se utiliza el estadístico de signo de la dispersión para construir y optimizar un gráfico de control que opera sobre una característica de calidad cuya medición está afectada por errores de redondeo debidos a la resolución del instrumento de medida.

Los errores en los sistemas de medida pueden ser de diferente índole. En concreto, según Linna y Woodall (2001), se pueden modelizar utilizando la ecuación (2.9):

$$X'_j = \left\lfloor \frac{A + BX_j + \varepsilon_j}{\rho} + \frac{1}{2} \right\rfloor \rho \quad (2.9)$$

donde X_j corresponde al valor real de la variable medida, X'_j representa el valor que nos devuelve el instrumento de medida, el par (A, B) representa el llamado error lineal de sesgo (*bias-linearity error*), ε_j corresponde al error de *precisión* que se cuantifica a través del ruido o variabilidad en la medida, y ρ es el error de *redondeo* que recoge la *resolución* del instrumento de medida. En la expresión superior aparece la función suelo $\lfloor \cdot \rfloor$, que consiste en redondear a la baja y sin decimales lo que se encuentra entre el símbolo “[” y el símbolo “]” (es decir, la función suelo devuelve un número entero).

Como este trabajo se centra en estudiar únicamente el impacto del error de redondeo ρ , la ecuación (2.9) queda modificada y se convierte en la ecuación (2.10):

$$X'_j = \left\lfloor \frac{X_j}{\rho} + \frac{1}{2} \right\rfloor \rho \quad (2.10)$$

Así, mediante esta expresión es posible convertir cualquier valor real en el valor observado, pues conociéndose la resolución ρ del instrumento de medida es posible calcular cuál será el valor que obtendremos en la medición.

2.4. Distribuciones de Johnson

2.4.1. Definición

Durante las experiencias numéricas que se llevan a cabo en este trabajo y cuyos resultados se muestran en el capítulo 4, se realizará la suposición de que la característica de calidad cuya dispersión se busca monitorizar posee una distribución de probabilidad que sigue un modelo de Johnson. Esta familia de distribuciones fue definida por primera vez por Johnson (1949). Se trata de una familia de distribuciones muy versátil, ya que puede adaptarse a valores de curtosis y asimetría muy distintos. Esto hace que, en la práctica, muchas características continuas con comportamiento aleatorio puedan ser ajustadas de manera bastante aproximada a una distribución de Johnson. De hecho, la flexibilidad que poseen estas distribuciones permite que tengan aplicaciones en campos muy variados; por ejemplo, en epidemiología (Flynn, 2006), en finanzas (Brooks et al., 2005) y en control estadístico de la calidad (Castagliola y García Castellanos, 2008).

Matemáticamente, se dice que una variable aleatoria X sigue una distribución de Johnson con parámetros γ, δ, ζ y λ (los cuatro parámetros son reales y $\lambda, \delta > 0$) si puede ser reducida a una variable que sigue una normal tipificada (media nula y varianza unitaria) Z mediante la ecuación (2.11):

$$Z = \gamma + \delta f\left(\frac{X - \zeta}{\lambda}\right) \quad (2.11)$$

donde f puede ser una de las tres funciones que se muestran en las ecuaciones (2.12), (2.13) y (2.14), y que dan lugar a cada una de las tres familias en las que se dividen las distribuciones de Johnson: *bounded* (S_B), *semibounded* (S_L) y *unbounded* (S_U), respectivamente.

$$f(y) = \ln\left(\frac{y}{1-y}\right); \quad 0 < y < 1; \quad \zeta < X < \zeta + \lambda \quad (2.12)$$

$$f(y) = \ln(y); \quad 0 < y < +\infty; \quad \lambda = 1; \quad \zeta < X < +\infty \quad (2.13)$$

$$f(y) = \sinh^{-1}(y); \quad -\infty < y < +\infty; \quad -\infty < X < +\infty \quad (2.14)$$

Además de las tres familias mencionadas, existe una cuarta familia llamada familia normal (S_N) donde $f(y) = y$. Por tanto, puede decirse que la distribución normal es un caso particular de las distribuciones de Johnson.

A partir de estas expresiones, se pueden obtener las funciones de distribución acumulada F_X (CDF, de las siglas en inglés *cumulative distribution function*) y la inversa de estas F_X^{-1} . A continuación, se muestran ambas funciones para la familia *bounded* (ecuaciones (2.15) y (2.16)), para la familia *semibounded* (ecuaciones (2.17) y (2.18)) y para la familia *unbounded* (ecuaciones (2.19) y (2.20)).

$$F_X(x) = \begin{cases} \Phi\left(\gamma + \delta \ln\left(\frac{x - \zeta}{\zeta + \lambda - x}\right)\right), & \zeta < x < \zeta + \lambda \\ 0, & x \leq \zeta \\ 1, & x \geq \zeta + \lambda \end{cases} \quad (2.15)$$

$$F_X^{-1}(u) = \zeta + \frac{\lambda}{1 + e^{\frac{\gamma - \Phi^{-1}(u)}{\delta}}}, \quad 0 < u < 1 \quad (2.16)$$

$$F_X(x) = \begin{cases} \Phi(\gamma + \delta \ln(x - \zeta)), & x > \zeta \\ 0, & x \leq \zeta \end{cases} \quad (2.17)$$

$$F_X^{-1}(u) = \zeta + e^{\frac{\Phi^{-1}(u) - \gamma}{\lambda}}, \quad 0 < u < 1 \quad (2.18)$$

$$F_X(x) = \Phi\left(\gamma + \delta \sinh^{-1}\left(\frac{x - \zeta}{\zeta + \lambda - x}\right)\right), \quad -\infty < x < +\infty \quad (2.19)$$

$$F_X^{-1}(u) = \zeta + \lambda \sinh\left(\frac{\Phi^{-1}(u) - \gamma}{\delta}\right), \quad 0 < u < 1 \quad (2.20)$$

En estas ecuaciones aparece la función Φ , la cual corresponde a la función de distribución acumulada de la normal tipificada. Además, hay que remarcar que se asume $\lambda = 1$ para el caso *semibounded*.

Sobre el significado de los parámetros, es importante destacar, a modo intuitivo, que ζ se relaciona con la localización de la variable X ; λ con su dispersión; y, finalmente, δ y γ son parámetros de forma.

2.4.2. Selección de distribuciones para un banco de pruebas

Una vez se ha explicado qué son las distribuciones de Johnson y qué parámetros las definen, se muestra en la tabla 1 una selección de 18 distribuciones de Johnson que serán utilizadas para los experimentos numéricos llevados a cabo en el capítulo 4. Esta selección es utilizada en el artículo de Castagliola et al. (2020), que sirve de referencia para el presente trabajo, como ya se ha mencionado en la introducción. Las distribuciones que forman parte de esta selección comparten, todas ellas, mediana igual a cero y desviación típica igual a 1. En cuanto a su asimetría ($\sqrt{\beta_1}$) y su curtosis ($\beta_2 - 3$), las 18 distribuciones cubren un amplio rango de posibles valores, lo cual las convierte en un conjunto representativo de las diferentes situaciones que en la práctica pueden producirse.

#	$\sqrt{\beta_1}$	$\beta_2 - 3$	Familia	γ	δ	ζ	λ
1	0	-1,2	Bounded	0	0,6465	-1,8153	3,6306
2	0	-0,6	Bounded	0	1,3983	-3,1097	6,2195
3	0	0	Unbounded	0	100	0	100
4	0	1	Unbounded	0	2,3212	0	2,1094
5	0	3	Unbounded	0	1,6104	0	1,3118
6	0	6	Unbounded	0	1,3493	0	1
7	2	4,3	Bounded	1,7464	0,6908	-0,4893	6,6213
8	2	6,1	Bounded	3,3279	1,2270	-1,0016	16,0883
9	2	7,9	Unbounded	-4,8560	1,8044	-1,4190	0,1933
10	2	10,8	Unbounded	-1,0444	1,4320	-0,6554	0,8236
11	2	16,7	Unbounded	-0,5298	1,2093	-0,3315	0,7331
12	2	25,5	Unbounded	-0,3437	1,0892	-0,2023	0,6305
13	5	39,9	Bounded	3,3715	0,7459	-0,2709	25,1500
14	5	52,6	Bounded	5,2193	0,9813	-0,4732	97,0433
15	5	65,3	Unbounded	-4,0187	1,0864	-0,5665	0,0281
16	5	86,4	Unbounded	-0,7570	0,9874	-0,3203	0,3795
17	5	128,7	Unbounded	-0,4319	0,9080	-0,1854	0,3754
18	5	192,1	Unbounded	-0,2987	0,8556	-0,1212	0,3403

Tabla 2.1. Parámetros y coeficientes para cada una de las dieciocho distribuciones de Johnson que se usarán en los experimentos computacionales del capítulo 4.

Nótese que la distribución normal (familia S_N) estaría asociada los valores de asimetría $\sqrt{\beta_1} = 0$ y curtosis $\beta_2 = 3$, la cual corresponde a la distribución de Johnson numerada con un 3. Sin embargo, puede también ser aproximada mediante una distribución de Johnson de la familia S_U con $\gamma = 0$ y $\delta \rightarrow \infty$. Por otro lado, la distribución de Johnson numerada con un 1, tiene un gran parecido a

la distribución uniforme, en términos de asimetría y curtosis (véase figura 2.1). Otras distribuciones que se asemejan a las distribuciones de Johnson que aparecen anteriormente son: la distribución triangular que se corresponde con la distribución de Johnson numerada con un 2; y las distribuciones t de Student con 10, 6 y 5 grados de libertad que se corresponden con las distribuciones de Johnson numeradas con un 4, 5 y 6 respectivamente (según Castagliola et al., 2020)

La figura 2.1 representa las funciones de densidad de las 18 distribuciones de Johnson consideradas en la tabla 2.1.

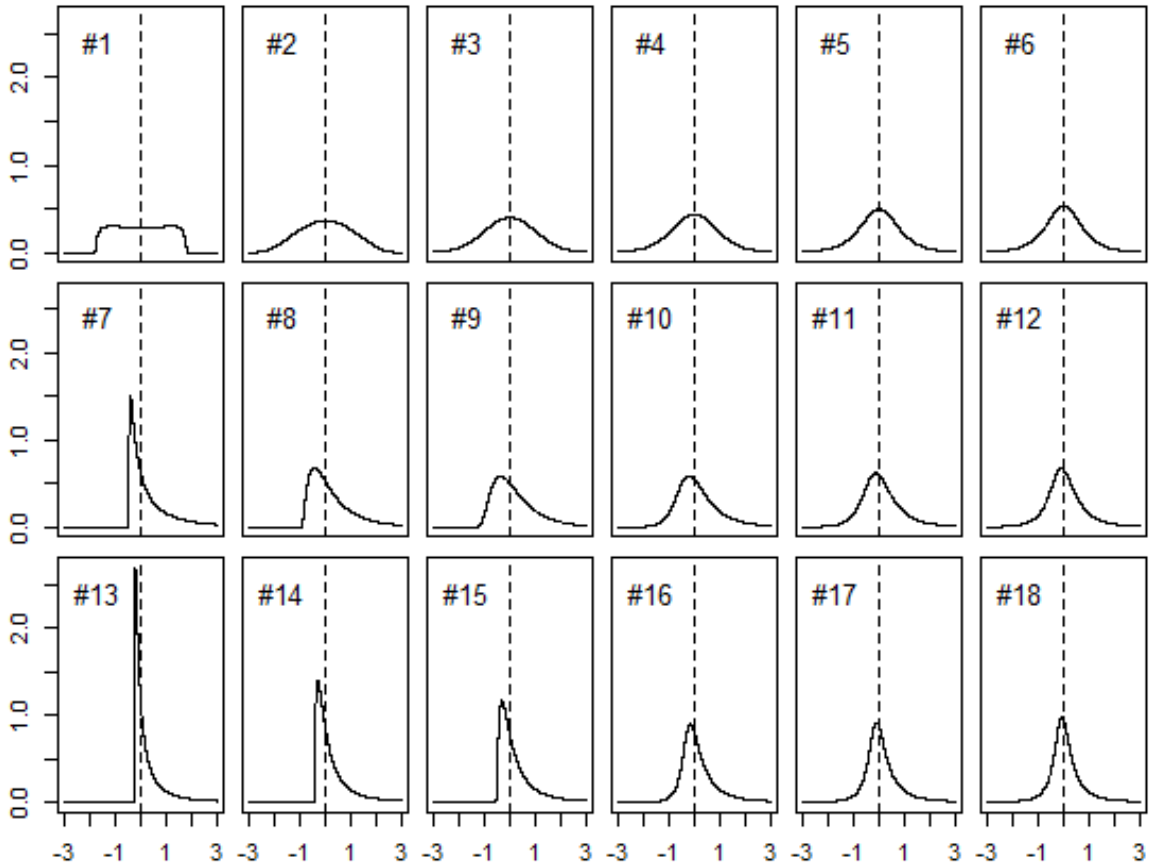


Figura 2.1. Gráficas de las dieciocho distribuciones de Johnson. Elaboración propia a partir de código de R de V. Giner.

Además de las gráficas mostradas anteriormente, se ubica a cada distribución en función de sus dos coeficientes $\sqrt{\beta_1}$ y β_2 en la figura 2.2, además de aclarar qué zonas del gráfico corresponde a qué familia. Nótese que, según Balakrishnan y Lai (2009), los coeficientes de asimetría y curtosis siempre cumplen la relación $\beta_1 \geq \beta_2 + 1$. La zona que no cumple esta relación es la que aparece indicada como “impossible zone” en la figura.

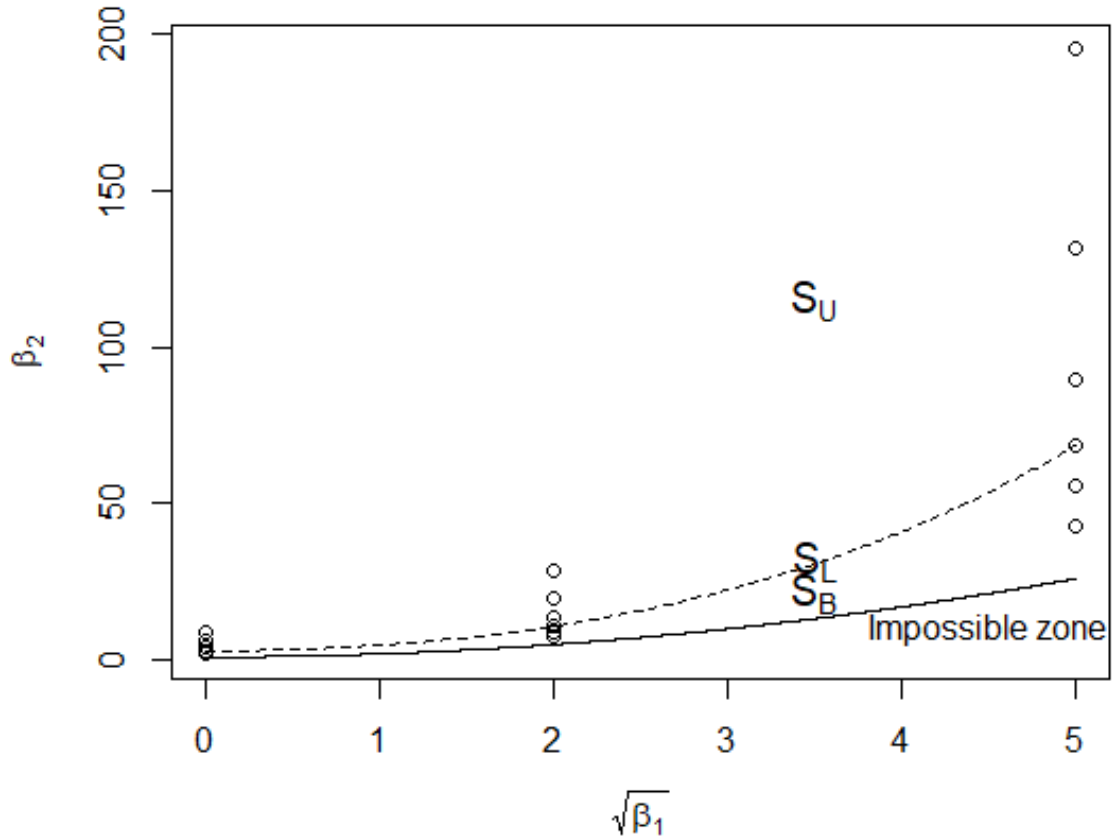


Figura 2.2. Gráfico en los que se ubica cada distribución de Johnson en función de sus dos coeficientes. Elaboración propia a partir de código de R de V. Giner.

2.4.3 Determinación de los parámetros de una distribución de Johnson con un cambio de dispersión τ

Durante este trabajo se menciona el parámetro τ que no es más que el cambio de dispersión que se pretende detectar. Es decir, $\tau = \sigma_1/\sigma_0$ donde σ_1 corresponde a la desviación que posee la variable aleatoria X cuando el proceso se encuentra fuera de control y σ_0 corresponde a la desviación que posee el proceso cuando se encuentra bajo control. De este modo, las distribuciones de Johnson se verán transformadas cuando el proceso se sale de control variando sus parámetros. Se puede demostrar que, si se tiene un proceso con mediana cero y se le modifica la desviación típica manteniendo la asimetría y la curtosis constantes, la distribución de Johnson con unos parámetros iniciales $(\gamma, \delta, \zeta, \lambda)$ se transforma en una distribución de Johnson $(\gamma, \delta, \tau\zeta, \tau\lambda)$.

3. Metodología

Tras haber presentado en el capítulo anterior los antecedentes y el conocimiento previo que es necesario tener en cuenta, en el presente capítulo desarrollamos nuestra propuesta metodológica.

Las tareas que se han llevado a cabo para conseguir los objetivos de este trabajo pueden agruparse en tres fases o etapas diferenciadas:

- Etapa 1: Diseñar un procedimiento para determinar los valores óptimos de los parámetros del gráfico de control no paramétrico tipo Shewhart para la dispersión basado en el estadístico de signo de Pawar et al. (2018), ignorando posibles errores de redondeo, ampliándolo para tratar también el caso en que sea de interés detectar una *disminución* de la variabilidad. Las condiciones iniciales o parámetros del problema que se considerarán como 'inputs' serán: n (tamaño de la muestra), τ (cambio en la dispersión) y J (distribución de Johnson); y mostrar la eficiencia de cada gráfico expresada en términos de ARL_0 y ARL_1 . Véanse más detalles sobre la notación en la sección 3.1.
- Etapa 2: Recalcular los valores reales de ARL_0 y ARL_1 cuando los gráficos optimizados en la etapa 1 se aplican en contextos en los que se producen errores de redondeo en el instrumento de medida (en la forma descrita en la sección 2.3). Esto permitirá evaluar y cuantificar la *degradación* de las propiedades de dichos gráficos.
- Etapa 3: *Reoptimizar* el gráfico objeto de estudio para que tenga en cuenta los errores de redondeo; esto es, diseñar un procedimiento para determinar los valores óptimos del gráfico no paramétrico propuesto en presencia de errores de redondeo.

A lo largo de este trabajo, nos referiremos a los resultados obtenidos tras las etapas 1, 2 y 3 como *situaciones 1, 2 y 3*, respectivamente.

A continuación, introducimos la notación que es común a las tres etapas. Posteriormente, se procede a detallar cuáles son las ecuaciones y los algoritmos desarrollados para llevar a cabo cada una de estas fases.

3.1. Notación

En esta sección, se busca exponer toda la notación común a las tres etapas, aunque parte de esta ya ha aparecido en el capítulo 2.

Notación básica:

- X : variable aleatoria que se está monitorizando.
- I_L y I_U : umbrales inferior y superior (véase ecuaciones (2.2) y (2.3)).

Parámetros del problema:

- α_0 : parámetro del problema que indica la probabilidad de error de primera especie tope (=0,27%, criterio tres-sigma).
- n : parámetro del problema que nos indica el tamaño muestral.
- τ : parámetro del problema que indica el cambio en la dispersión calculado como σ_1/σ_0 (véase sección 2.4.3).
- J : parámetro del problema que indica la distribución de Johnson subyacente para X (véase sección 2.4.2).
- ρ : parámetro del problema que indica la resolución del instrumento de medida (solo aparece en las etapas 2 y 3).

Parámetros del modelo:

- p_0 : parámetro del modelo que indica la probabilidad de que una medición de X quede fuera de los umbrales I_L y I_U .
- L : límite de control inferior si $\tau < 1$ o límite de control superior si $\tau > 1$.

Medidas de desempeño:

- α : probabilidad de error de primera especie o de falsa alarma.
- β : probabilidad de error de segunda especie.
- ARL_0 : longitud media de racha para el proceso bajo control (véase ecuación (1.1)).
- ARL_1 : longitud media de racha para el proceso fuera de control (véase ecuación (1.2)).

3.2. Etapa 1: diseño óptimo del gráfico de control para la dispersión para mediciones sin errores de redondeo

Para la primera parte de este capítulo se busca diseñar de manera óptima un gráfico de control para la dispersión basado en el estadístico de signo y sin tener en cuenta los posibles errores derivados del instrumento de medida. Para ello es necesario hacer uso de las ecuaciones (2.2) y (2.3), ya explicadas, que nos definen el umbral inferior I_L y el umbral superior I_U , y la relación de estos umbrales con p_0 . Además, se hace uso de la definición del estadístico de signo U_t , que corresponde a la ecuación (2.1).

A partir de estas ecuaciones y teniendo en cuenta que el estadístico de signo sigue una distribución relacionada con la binomial en el sentido en que aparece en la ecuación (2.5), se encuentran las expresiones para la probabilidad de error de primera especie, α , según tengamos un valor de $\tau < 1$ obteniéndose la ecuación (3.1) (es decir, se buscan decrementos significativos en la dispersión) o según

tengamos un valor de $\tau > 1$, obteniéndose la ecuación (3.2) (es decir, se buscan incrementos significativos en la dispersión).

$$\alpha = \Pr(U_t < LCL) | \sigma = \sigma_0) = \sum_{i=0}^{\frac{LCL+n}{2}-1} \Pr(V_t = i | p = p_0) \quad (3.1)$$

$$\alpha = \Pr(U_t > UCL) | \sigma = \sigma_0) = \sum_{\frac{UCL+n}{2}+1}^n \Pr(V_t = i | p = p_0) \quad (3.2)$$

Para obtener la probabilidad de error de segunda especie, β , se tiene la ecuación (3.3) si $\tau < 1$, y la ecuación (3.4) si $\tau > 1$. En estas ecuaciones se tiene que $\tau > 0$ y que $\tau \neq 1$.

$$\beta(\tau) = \Pr(U_t \geq LCL) | \sigma = \sigma_1) = \sum_{i=\frac{LCL+n}{2}}^n \Pr(V_t = i | p = p_1) \quad (3.3)$$

$$\beta(\tau) = \Pr(U_t \leq UCL) | \sigma = \sigma_1) = \sum_{i=0}^{\frac{UCL+n}{2}} \Pr(V_t = i | p = p_1) \quad (3.4)$$

En estas ecuaciones conviene recordar:

$$\sigma_1 = \tau \sigma_0 \quad (3.5)$$

$$V_t \sim Bi(n, p_0) \quad (3.6)$$

donde σ_0 corresponde a la desviación típica de la variable aleatoria X cuando el proceso se encuentra bajo control, σ_1 corresponde a la desviación típica que se pretende detectar y τ es la variación de la dispersión que se pretende detectar.

Además, respecto a p_0 y p_1 es importante matizar que:

$$p_0 = \Pr(X \notin [L_L, I_U] | \sigma = \sigma_0) \quad (3.7)$$

$$p_1 = \Pr(X \notin [L_L, I_U] | \sigma = \sigma_1) \quad (3.8)$$

Al obtener tanto α como $\beta(\tau)$ se pueden obtener tanto $ARL_0(\alpha)$ como $ARL_1(\beta)$ mediante las ecuaciones (1.1) y (1.2) respectivamente.

Una vez se ha explicado cómo obtener α y β para todas las situaciones se procede a detallar como se debe realizar la optimización.

El algoritmo de optimización procesará diferentes situaciones iniciales que se resumen en tres factores: el tamaño de muestra n que toma los siguientes valores $\{10, 15, 20, 25, 30\}$; el cambio de dispersión a detectar τ que toma los siguientes valores $\{0.25, 0.5, 2, 4\}$; y, finalmente, la distribución de Johnson subyacente a la variable aleatoria X, J , (véase sección 2.4) que toma los siguientes valores $\{1, 2, 3, \dots, 16, 17, 18\}$. Este será nuestro *benchmark* o banco de pruebas para las experiencias numéricas del capítulo 4; desde el punto de vista del diseño de experimentos, puede ser visto como un diseño factorial completo.

Para encontrar el gráfico óptimo, el algoritmo debe encontrar el valor óptimo de p_0 que se denota como p_0^* y el valor óptimo de LCL o UCL según sea $\tau < 1$ o $\tau > 1$. En el algoritmo se busca el valor óptimo de LCL o UCL mediante la variable de decisión L que corresponderá a un u otro caso según el valor de τ (el valor óptimo de L se denota como L^*)

Así, la optimización que se busca llevar a cabo es:

$$\begin{cases} \text{Min} & \beta(p_0, L|n, \tau) \\ \text{s. a:} & \alpha(p_0, L|n) \leq \alpha_0, \\ & p_0 \in R, \quad L \in S \end{cases} \quad (3.9)$$

donde el conjunto $R = \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95\}$, donde el conjunto $S = \{-n, -n + 2, \dots, n - 2, n\}$ y donde $\alpha_0 = 0.0027$ si se busca que, como mínimo, el proceso siga el criterio tres-sigma; o bien cualquier otro valor de interés para el usuario final.

De este modo, se busca minimizar la probabilidad de error de segunda especie (por lo que se está minimizando también ARL_1), sin que esta minimización conlleve renunciar a que el proceso tenga un ARL_0 descontrolado.

Usualmente, se utiliza la restricción $\alpha(p_0, L|n) = \alpha_0$, pero, como la optimización se realiza sobre valores discretos, no se puede garantizar el cumplimiento de la igualdad anterior; por ello, se vuelve necesario recurrir a limitar un tope para la probabilidad de error de primera especie en vez de tratar de igualarla con este tope. De este modo, se busca un valor de ARL_0 no más pequeño que 370,4, en el caso de seguir el criterio tres-sigma, y un ARL_1 que sea mínimo, cumpliendo esta restricción.

Aunque en nuestro caso tomaremos valores discretos, p_0 es una variable continua que se sitúa entre 0 y 1 por lo que realmente no se puede asegurar una optimalidad real en el sentido de que el p_0 óptimo podría no ser uno de los valores discretos escogidos. Es por esto por lo que nos enfrentamos a un problema de optimización discreta no lineal.

A continuación, se muestra en pseudocódigo (algoritmo 1) el método que soluciona el problema caracterizado en la ecuación (3.9). Este algoritmo se ha desarrollado a partir de una idea previa no publicada de V. Giner y P. Castagliola. El algoritmo aplica una estrategia enumerativa implícita. La estrategia es enumerativa porque se revisan y se cubren todos los posibles valores de p_0 y L (contenidos en el conjunto R y en el conjunto S , respectivamente), tal y como se observa en los dos bucles *for*. Por otro lado, la estrategia es implícita porque se descartan algunos valores de L mediante un *break* cuando no se cumple la única restricción impuesta: $\alpha(p_0, L|n) \leq \alpha_0$.

En el capítulo 4, se muestran los resultados numéricos obtenidos al optimizar el gráfico en esta etapa 1. A estos resultados los llamaremos “Situación 1”.

ALGORITMO 1: Diseño óptimo del gráfico de control de Shewhart no paramétrico para la dispersión basado en el estadístico de signo sin errores de redondeo

Input: n, τ, J, α_0 .

Output: Valores óptimos de p_0, L, β, α y p_1 .

// F_X^{-1} corresponde a la inversa de la cdf de $X \sim J$.

// U_t corresponde al estadístico de signo (ecuación (2.1)).

// $V_t \sim Bi(p_0, n)$

// Inicializar variables

$p_0^*, L^* \leftarrow null$

$\beta \leftarrow +\infty$

$\alpha \leftarrow null$

$\sigma_1 \leftarrow \tau\sigma_0$

// Se realizan dos bucles para visitar todas las soluciones posibles

foreach $p_0 \in R$ **do**

$I_L \leftarrow F_X^{-1}(p_0/2 \mid \sigma = \sigma_0)$

Ecuación (1.2)

$I_U \leftarrow F_X^{-1}(1 - p_0/2 \mid \sigma = \sigma_0)$

Ecuación (1.3)

$p_1 \leftarrow Pr(X \notin [I_L, I_U] \mid \sigma = \sigma_1)$

Ecuación (3.8)

for $L \leftarrow n$ **to** $-n$ **to step** -2 **do**

if $L = n$ **then**

$\alpha \leftarrow 0$

$\beta \leftarrow 1$

else

if $\tau < 1$ **then**

$\alpha \leftarrow \alpha + Pr(V_t = ((-L - 2) + n)/2 \mid p = p_0)$

Ecuación (3.1)

else

$\alpha \leftarrow \alpha + Pr(V_t = ((L + 2) + n)/2 \mid p = p_0)$

Ecuación (3.2)

end if

if $\alpha > \alpha_0$ **then break**

if $\tau < 1$ **then**

$\beta \leftarrow \beta - Pr(V_t = ((-L - 2) + n)/2 \mid p = p_1)$

Ecuación (3.3)

else

$\beta \leftarrow \beta - Pr(V_t = ((L + 2) + n)/2 \mid p = p_1)$

Ecuación (3.4)

end if

end if

// Si llegamos aquí, los actuales (p_0 y L) son una posible solución.

if $\beta < \beta^*$ **then**

$p_0^* \leftarrow p_0$

if $\tau < 1$ **then**

$L^* \leftarrow -L$

else

$L^* \leftarrow +L$

end if

$\beta^* \leftarrow \beta$

$\alpha^* \leftarrow \alpha$

$p_1^* \leftarrow p_1$

end if

end for

end foreach

return Solución óptima: (p_0^*, L^*) . β óptima: β^* . α real: α^*

3.3. Etapa 2: cálculo del ARL_0 y ARL_1 real ante la presencia de *empates* en los gráficos diseñados

En esta sección se busca calcular cual serían los valores reales tanto de ARL_0 como de ARL_1 si al utilizar los gráficos obtenidos en la sección 3.1 apareciera un error de redondeo ρ tal y como se explica en la sección 2.3.

Para esto, en primer lugar, es necesario definir la probabilidad de que la medición nos devuelva cada uno de los tres valores que puede tomar el estadístico de signo (véase ecuación (2.1)). Estas tres probabilidades corresponden a las ecuaciones (3.10), (3.11) y (3.12).

De este modo, la probabilidad de que la medición de la variable X quede fuera de $[I_L, I_U]$ es:

$$\pi_1 = \Pr(X' < I_L) + \Pr(X' > I_U) = F_X\left(I_L - \frac{\rho}{2}\right) + 1 - F_X\left(I_U + \frac{\rho}{2}\right) \quad (3.10)$$

Mientras que la probabilidad de que la medición de la variable X quede dentro de los umbrales es:

$$\pi_{-1} = \Pr(I_L < X' < I_U)$$

$$\pi_{-1} = \begin{cases} F_X\left(I_U - \frac{\rho}{2}\right) - F_X\left(I_L + \frac{\rho}{2}\right), & \text{si } I_L + \frac{\rho}{2} \leq I_U - \frac{\rho}{2} \\ 0, & \text{si } I_L + \frac{\rho}{2} > I_U - \frac{\rho}{2} \end{cases} \quad (3.11)$$

Esto se debe a que la zona donde se pueden producir empates en el umbral inferior se puede solapar con la zona donde hay un empate para el umbral superior como se muestra en la figura 3.1 (en la figura 3.1 se muestra como ejemplo con $p_0 = 0.8$, $L = -1$, $\tau = 0.25$, $J = 13$ y $\rho = 0.2$). En la figura 3.1, la primera línea vertical y la cuarta línea vertical delimitan la zona de empate entorno al umbral inferior cuya posición la marca la segunda línea vertical, mientras que la tercera línea vertical y la sexta línea vertical delimitan la zona de empate entorno al umbral superior cuya posición la marca la quinta línea vertical.

De este modo, entre la tercera línea vertical y la cuarta línea vertical, se produce un solapamiento entre las zonas de empate por lo que la probabilidad de que la medición se sitúe entre los dos umbrales es nula.

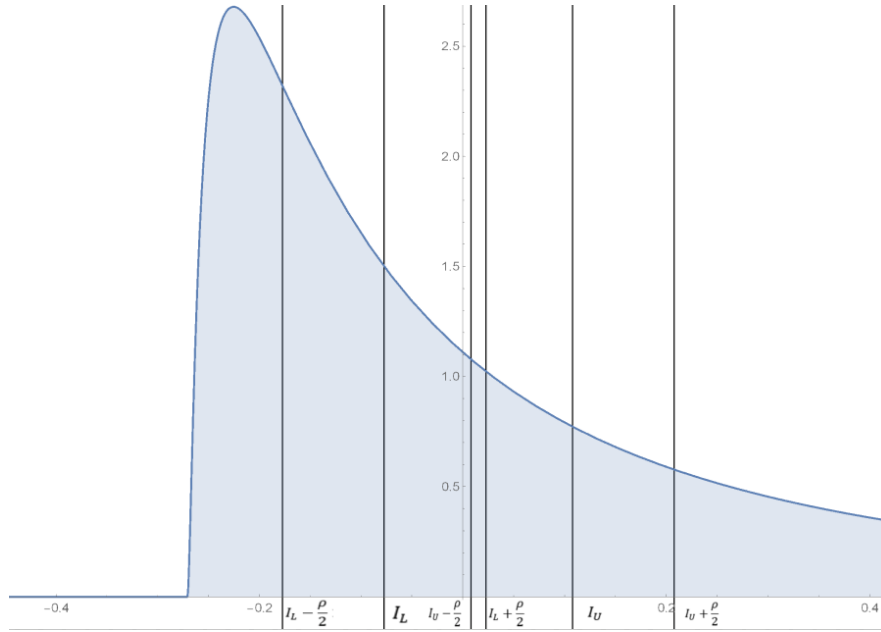


Figura 3.1. Gráfica que sirve de ejemplo para mostrar el solapamiento entre zonas de empate.

Finalmente, la probabilidad de que la medición se sitúe en I_L o I_U es:

$$\pi_0 = \Pr(X' = I_L) + \Pr(X' = I_U) = 1 - \pi_1 - \pi_{-1} \quad (3.12)$$

Esto se da ya que el error de redondeo no permite discriminar si el valor real se encuentra a la derecha o a la izquierda de I_L o I_U . A este fenómeno se le llama empate (“tie” en inglés) según Castagliola et al. (2020).

Nótese que, en el caso de que no hubiera empates ($\rho = 0$), tendríamos $\pi_1 = p_0$, $\pi_{-1} = 1 - p_0$ y $\pi_0 = 0$.

De esta forma, para calcular las tres probabilidades se pueden utilizar las ecuaciones (3.10), (3.11) y (3.12), además de calcular I_L y I_U mediante las ecuaciones (2.2) y (2.3). Además, hay que tener en cuenta que, según busquemos conocer las tres probabilidades en un proceso bajo control o fuera de control se deben modificar los parámetros de la distribución de Johnson $(\gamma, \delta, \zeta, \lambda)$ transformándolos en una distribución de Johnson $(\gamma, \delta, \tau\zeta, \tau\lambda)$ tal y como se explicó al final de la sección 2.4.

Una vez se tienen las tres probabilidades calculadas, queda calcular cual es la probabilidad de que dado un tamaño de muestra n , una distribución de Johnson J , una resolución del instrumento de medida ρ , un cambio en la dispersión τ , un límite de control L y una probabilidad p_0 , se consiga aceptar U_t como válido. Para esto se utiliza la ecuación (3.13) que nos indica cuál es la probabilidad de que U_t sea un valor concreto u .

$$\Pr(U_t = u) = \sum_{i=\max(0,-u)}^{\lfloor \frac{n-u}{2} \rfloor} \binom{n}{i} \binom{n-i}{u+i} \pi_{-1}^i \pi_0^{n-u-2i} \pi_{+1}^{u+i} \quad (3.13)$$

Así, para conocer cuál es la probabilidad de que se acepte un U_t solo hace falta realizar la suma que aparece en la ecuación (2.4).

$$\Pr(U_t \in [LCL, UCL]) = \sum_{i=LCL}^{UCL} \Pr(U_t = i) \quad (3.14)$$

En la ecuación (3.14), si se cumple que $\tau < 1$ se tendrá que $LCL = L$ y $UCL = n$ y si no, $LCL = -n$ y $UCL = L$.

Con estas ecuaciones ya se pueden obtener los valores reales de α y β pues se consiguen mediante las ecuaciones (3.15) y (3.16).

$$\alpha = 1 - \Pr(U_t \in [LCL, UCL] \mid \tau = 1) \quad (3.15)$$

$$\beta = \Pr(U_t \in [LCL, UCL] \mid \tau \neq 1) \quad (3.16)$$

En el caso en que $\tau = 1$ se trabaja con una distribución de Johnson de parámetros $(\gamma, \delta, \zeta, \lambda)$ mientras que para $\tau \neq 1$ se trabaja con una distribución de Johnson de parámetros $(\gamma, \delta, \tau\zeta, \tau\lambda)$.

Finalmente, los ARL reales se obtienen al utilizar el valor de α obtenido en la ecuación (3.15) y el valor de β obtenido en la ecuación (3.16) y mediante las ecuaciones (1.1) y (1.2).

3.4. Etapa 3: diseño óptimo del gráfico de control para la dispersión en presencia de empates

Para diseñar de manera óptima un gráfico de control para la dispersión basado en el estadístico de signo que tenga en cuenta la presencia de errores de redondeo resulta necesario modificar el algoritmo empleado en la sección 3.1, añadiendo algunas de las ecuaciones ya vistas en la sección 3.2.

A continuación (algoritmo 2), se expone el nuevo algoritmo en pseudocódigo, resultante de adaptar y modificar pertinentemente el algoritmo 1. Concretamente, se deja de utilizar p_1 y se calculan en su lugar, las probabilidades π_1 (probabilidad de rechazar), π_0 (probabilidad de empate) y π_{-1} (probabilidad de aceptar) correctamente (mediante las ecuaciones (3.10), (3.11) y (3.12)), al igual que la nueva función de masa de probabilidad para obtener los valores de α y β (mediante las ecuaciones (3.13) y (3.14)).

ALGORITMO 2: Diseño óptimo del gráfico de control de Shewhart no paramétrico para la dispersión basado en el estadístico de signo con errores de redondeo

Input: $n, \tau, J, \rho, \alpha_0$.

Output: Valores óptimos de p_0, L, β, α y p_1 .

// F_X corresponde a la cdf de $X \sim J$.

// F_X^{-1} corresponde a la inversa de la cdf de $X \sim J$.

// U_t corresponde al estadístico de signo (ecuación (2.1)).

// $V_t \sim Bi(p_0, n)$

// Inicializar variables

$p_0^*, L^* \leftarrow null$

$\beta \leftarrow +\infty$

$\alpha \leftarrow null$

$\sigma_1 \leftarrow \tau\sigma_0$

// Se realizan dos bucles para visitar todas las soluciones posibles

foreach $p_0 \in R$ **do**

$I_L \leftarrow F_X^{-1}(p_0/2 \mid \sigma = \sigma_0)$ Ecuación (1.2)

$I_U \leftarrow F_X^{-1}(1 - p_0/2 \mid \sigma = \sigma_0)$ Ecuación (1.3)

$\pi_1(\tau = 1) \leftarrow F_X\left(I_L - \frac{\rho}{2} \mid \sigma = \sigma_0\right) + 1 - F_X\left(I_U + \frac{\rho}{2} \mid \sigma = \sigma_0\right)$ Ecuación (3.10)

if $I_L + \frac{\rho}{2} \leq I_U - \frac{\rho}{2}$ **then**

$\pi_{-1}(\tau = 1) \leftarrow F_X\left(I_U - \frac{\rho}{2} \mid \sigma = \sigma_0\right) - F_X\left(I_L + \frac{\rho}{2} \mid \sigma = \sigma_0\right)$ Ecuación (3.11)

else

$\pi_{-1}(\tau = 1) \leftarrow 0$

end if

$\pi_0(\tau = 1) \leftarrow 1 - \pi_1(\tau = 1) - \pi_{-1}(\tau = 1)$ Ecuación (3.12)

$\pi_1(\tau \neq 1) \leftarrow F_X\left(I_L - \frac{\rho}{2} \mid \sigma = \sigma_1\right) + 1 - F_X\left(I_U + \frac{\rho}{2} \mid \sigma = \sigma_1\right)$ Ecuación (3.10)

if $I_L + \frac{\rho}{2} \leq I_U - \frac{\rho}{2}$ **then**

$\pi_{-1}(\tau \neq 1) \leftarrow F_X\left(I_U - \frac{\rho}{2} \mid \sigma = \sigma_1\right) - F_X\left(I_L + \frac{\rho}{2} \mid \sigma = \sigma_1\right)$ Ecuación (3.11)

else

$\pi_{-1}(\tau \neq 1) \leftarrow 0$

end if

$\pi_0(\tau \neq 1) \leftarrow 1 - \pi_1(\tau \neq 1) - \pi_{-1}(\tau \neq 1)$ Ecuación (3.12)

for $L \leftarrow n$ **to** $-n$ **to step** -1 **do**

if $L = n$ **then**

$\alpha \leftarrow 0$

$\beta \leftarrow 1$

else

if $\tau < 1$ **then**

$\alpha \leftarrow \alpha + Pr(U_t = -L - 1 \mid \pi_1(\tau = 1), \pi_0(\tau = 1), \pi_{-1}(\tau = 1))$ Ecuación (3.13)

else

$\alpha \leftarrow \alpha + Pr(U_t = L + 1 \mid \pi_1(\tau = 1), \pi_0(\tau = 1), \pi_{-1}(\tau = 1))$ Ecuación (3.13)

end if

if $\alpha > \alpha_0$ **then break**

if $\tau < 1$ **then**
 $\beta \leftarrow \beta - Pr(U_t = -L - 1 | \pi_1(\tau \neq 1), \pi_0(\tau \neq 1), \pi_{-1}(\tau \neq 1))$ Ecuación (3.13)

else
 $\beta \leftarrow \beta - Pr(U_t = L + 1 | \pi_1(\tau \neq 1), \pi_0(\tau \neq 1), \pi_{-1}(\tau \neq 1))$ Ecuación (3.13)

end if

end if

// Si llegamos aquí, los actuales (p_0 y L) son una posible solución.

if $\beta < \beta^*$ **then**

$p_0^* \leftarrow p_0$

if $\tau < 1$ **then**

$L^* \leftarrow -L$

else

$L^* \leftarrow +L$

end if

$\beta^* \leftarrow \beta$

$\alpha^* \leftarrow \alpha$

end if

end for

end foreach

return Solución óptima: (p_0^*, L^*) . β óptima: β^* . α real: α^*

4. Resultados

En este capítulo, se muestran los resultados obtenidos al procesar las tres situaciones descritas en el capítulo 3 para todos los niveles de los cuatro factores n, τ, J y ρ (estos factores se hallan descritos en la sección 3.1), es decir, se ha procesado el banco de pruebas entero. Concretamente, ρ tomará los siguientes valores $\{0.05, 0.1, 0.2\}$.

De esta forma, se han utilizado las ecuaciones que aparecen en el capítulo 3 pero también el algoritmo 1 (véase sección 3.1) y el algoritmo 2 (véase sección 3.3). Estos algoritmos se han ejecutado utilizando la versión 6.1.0 del software Scilab. Las ejecuciones se han realizado en un ordenador Medion PC con el sistema operativo Windows 10 Home a 64 bits, con un procesador Intel Core i3-6006U con dos núcleos, una velocidad de 2.00 GHz y provisto de 7,89 Gb de memoria RAM utilizable.

Para acceder al código en Scilab, se proporciona en el anexo A una dirección web que da acceso al archivo .sce utilizado.

Una vez introducido el capítulo, se procede a analizar cada una de las tres situaciones en las secciones 4.1, 4.2 y 4.3, para poder después compararlas entre ellas en las secciones 4.4, 4.5, 4.6 y 4.7.

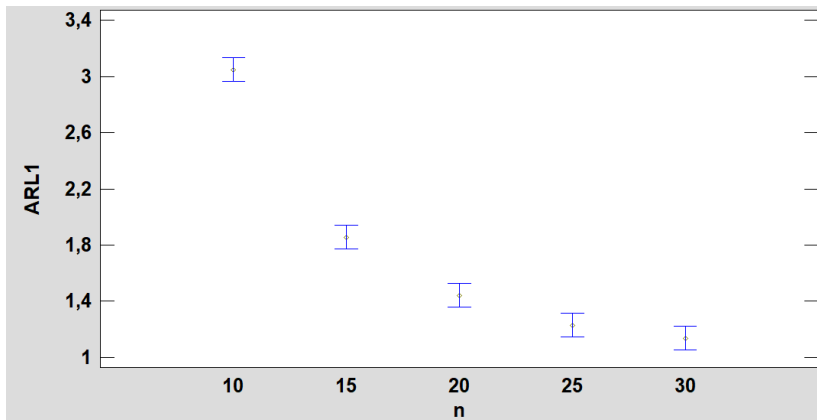
4.1. Análisis de los resultados obtenidos para la situación 1

En la situación 1, se busca diseñar de forma óptima un gráfico de control no paramétrico para la dispersión basado en el estadístico de signo en ausencia de empates (véase sección 3.1).

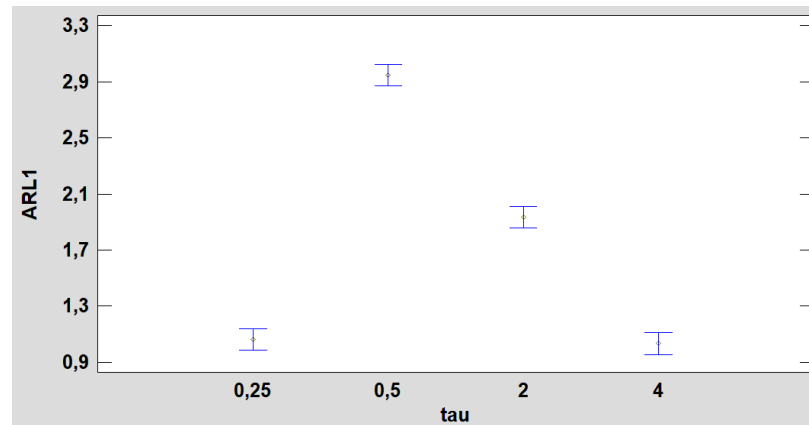
Al ejecutar el algoritmo 1, se han obtenido diferentes valores de ARL_0 y ARL_1 para cada una de las diferentes pruebas del banco de pruebas donde se combinan los siguientes tres factores: n (tamaño de muestra), τ (cambio en la dispersión) y J (distribución de Johnson). Estos resultados numéricos aparecen en la segunda hoja del Excel que se proporciona en el anexo B (la hoja se llama "Situación 1").

Con el fin de interpretar los resultados obtenidos, se ha realizado un análisis de la varianza (ANOVA) a fin de comprender, de modo descriptivo, como afecta cada uno de los tres factores, mencionados en el párrafo anterior, al ARL_1 (vease sección 1.2). La figura 4.1 muestra los gráficos de medias obtenidos al realizar el ANOVA con unos intervalos LSD al 95%.

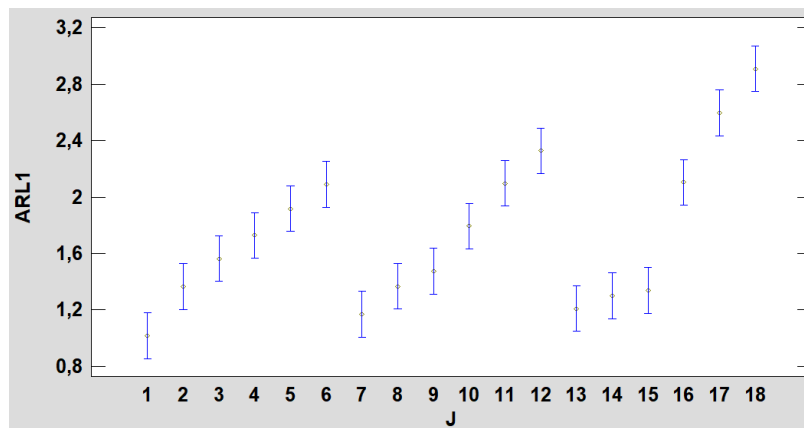
CAPÍTULO 4. RESULTADOS



(a) Efectos del tamaño de muestra, n , sobre el ARL_1



(b) Efectos del cambio de dispersión, τ , sobre el ARL_1



(c) Efectos de la distribución de Johnson, J , sobre el ARL_1

Figura 4.1 Gráficos de medias e intervalos LSD al 95% para los efectos del ANOVA en la situación 1.

La tabla del ANOVA se puede consultar en el anexo C, pues concretamente se encuentra titulada como tabla C.1. De hecho, es en esta tabla donde se observa que tanto los efectos principales como las interacciones de orden 2 son significativas (no se han buscado interacciones de mayor orden que 2).

En la figura 4.1 se aclara como afecta cada factor al ARL_1 , es decir, cual es la eficacia del gráfico para detectar que el proceso se ha salido de control. Así, para el factor n , tamaño de muestra, se halla que, a mayor tamaño de muestra, menor ARL_1 se tiene y por lo tanto mejor funciona el gráfico de control. Por otro lado, para el factor τ , cambio en la dispersión, se observa que para valores de τ que más disten de 1, menor es el ARL_1 , pues más fácil resulta detectar cambios grandes en la dispersión que cambios pequeños. Finalmente, para el factor J , distribución de Johnson, se obtiene que, para valores iguales de asimetría, cuanto mayor es la curtosis mayor es el ARL_1 y, por lo tanto, más difícil resulta detectar que el proceso se ha salido de control. Las tres conclusiones que arrojan las gráficas son altamente intuitivas.

Finalmente, en el anexo D, se ha construido una tabla de medias, concretamente la tabla D.1, que proporciona el valor promedio de ARL_0 y el valor promedio de ARL_1 para cada nivel de cada factor.

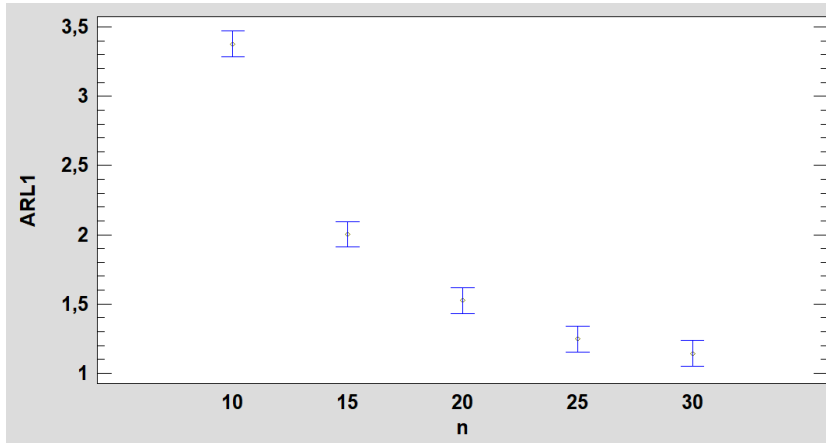
4.2. Análisis de los resultados obtenidos para la situación 2

En la situación 2, se busca recalcular los valores reales de ARL_0 y ARL_1 para cada una de las pruebas del banco de pruebas al utilizar el gráfico de control diseñado anteriormente, en la sección 4.1, para mediciones sin errores de redondeo o empates, cuando sí hay presencia de errores de redondeo.

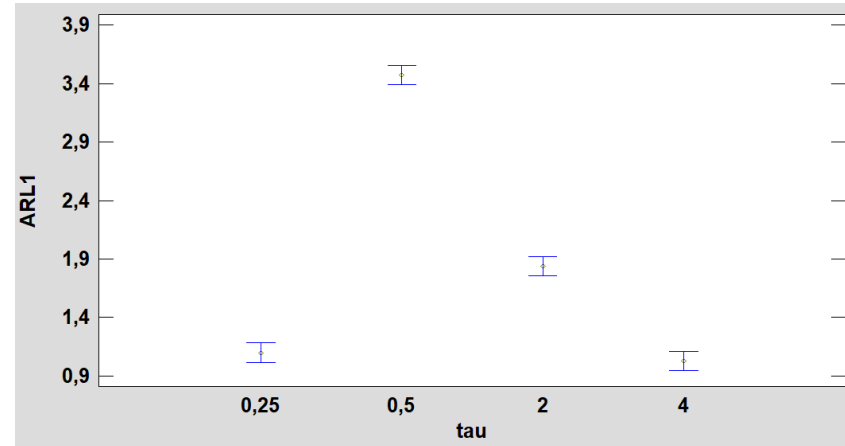
Los valores reales de ARL_0 y ARL_1 se consiguen recalcular al aplicar las ecuaciones (3.10), (3.11), (3.12), (3.13), (3.14), (3.15) y (3.16), es decir, hace falta aplicar todas las ecuaciones que se detallan en la sección 3.2.

Los resultados numéricos de esta sección se pueden consultar en el anexo B, donde se proporciona un enlace para acceder a un archivo Excel que contiene resultados numéricos. Concretamente, los valores correspondientes a esta sección se encuentran en la tercera hoja del Excel, llamada "Situación 2".

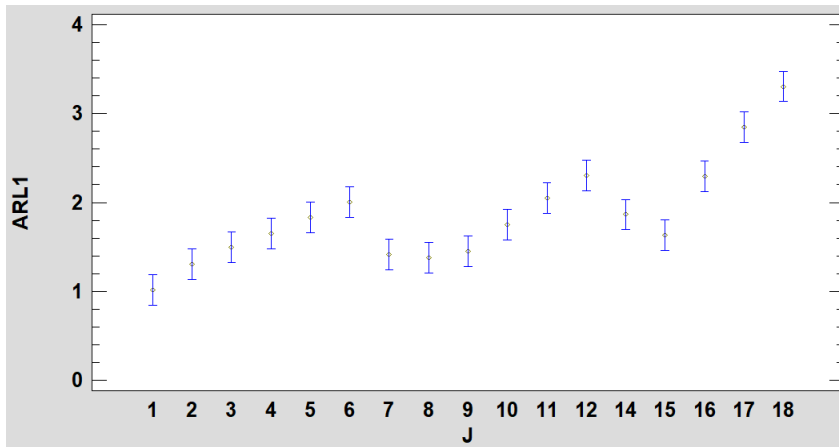
Con el fin de interpretar los resultados obtenidos, se ha realizado un análisis de varianza (ANOVA) que nos permite visualizar como afecta cada uno de los cuatro factores al ARL_1 (véase sección 1.2). La distribución $J = 13$ proporciona unos valores de ARL_1 muy extremos debido a la forma puntiaguda que posee (véase figura 2.1), por lo que se ha extraído del análisis para no distorsionar el ANOVA y la tabla de medias (anexo D). La figura 4.2 muestra los gráficos de medias obtenidos al realizar el ANOVA con unos intervalos LSD al 95%.



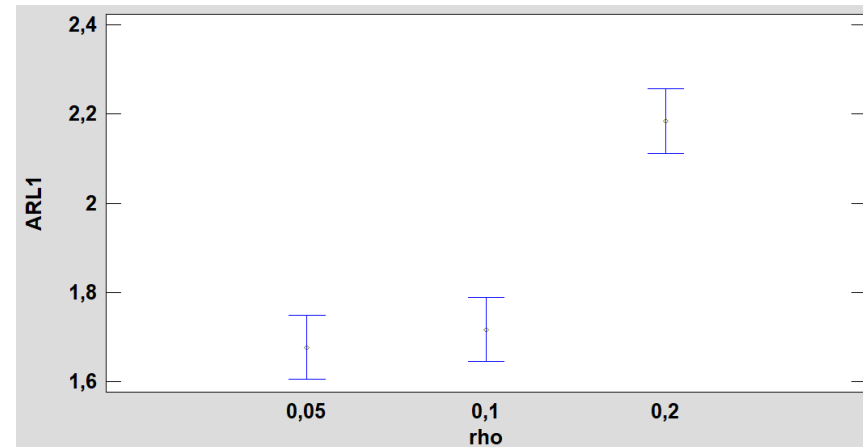
(a) Efectos del tamaño de muestra, n , sobre el ARL_1



(b) Efectos del cambio de dispersión, τ , sobre el ARL_1



(c) Efectos de la distribución de Johnson, J , sobre el ARL_1



(d) Efectos del error de redondeo, ρ , sobre el ARL_1

Figura 4.2 Gráficos de medias e intervalos LSD al 95% para los efectos del ANOVA en la situación 2.

La tabla del ANOVA se puede consultar en el anexo C (al igual que sucedía con el ANOVA de la sección anterior, es decir, la sección 4.1) y esta tabla se numera como tabla C.2. En esta tabla se observa cómo no solo los efectos principales son significativos, sino que también lo son las interacciones de orden 2.

En la figura 4.2, nos encontramos con el mismo patrón para los factores n, τ y J , que ya se observaba en la sección 4.1, concretamente en la figura 4.1. De este modo, para el factor n , tamaño de muestra, vemos que, a mayor tamaño de muestra, menor ARL_1 ; para el factor τ , cambio en la dispersión, se observa que para valores de τ que más disten de 1, menor es el ARL_1 ; para el factor J , distribución de Johnson, se obtiene que, para valores iguales de asimetría, cuanto mayor es la curtosis mayor es el ARL_1 . La diferencia con la figura 4.1 radica en la gráfica (d) de la figura 4.2 que nos muestra como afecta el error de redondeo ρ al ARL_1 , observándose que cuanto mayor es el error de redondeo más difícil resulta percatarse de que el proceso se ha salido de control, lo que hace aumentar ARL_1 .

Por último, en el anexo D, se ha construido una tabla de medias, la tabla D.1, que nos entrega el valor promedio de ARL_0 y el valor promedio de ARL_1 para cada nivel de cada factor.

4.3. Análisis de los resultados obtenidos para la situación 3

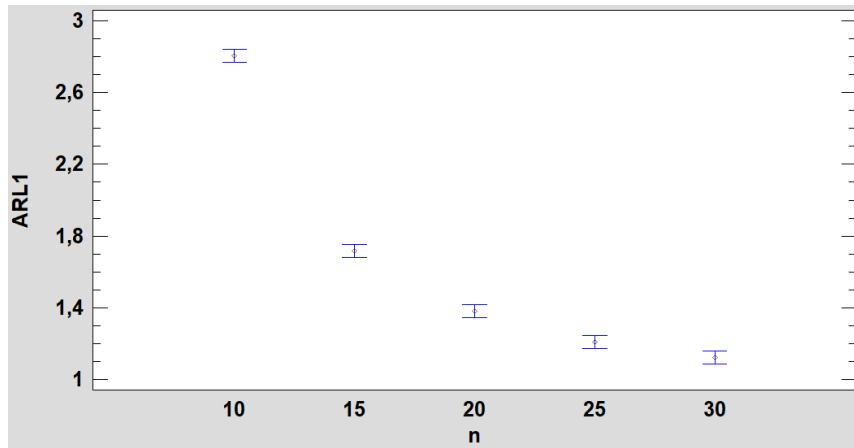
En la situación 3, se trata de diseñar de forma óptima un gráfico de control que tenga en cuenta la presencia de los errores de redondeo y poder, así, procesar el banco de pruebas.

Para ello, se hace uso del software Scilab para ejecutar el algoritmo 2 que aparece en la sección 3.3. El código en Scilab utilizado se encuentra disponible en el anexo A.

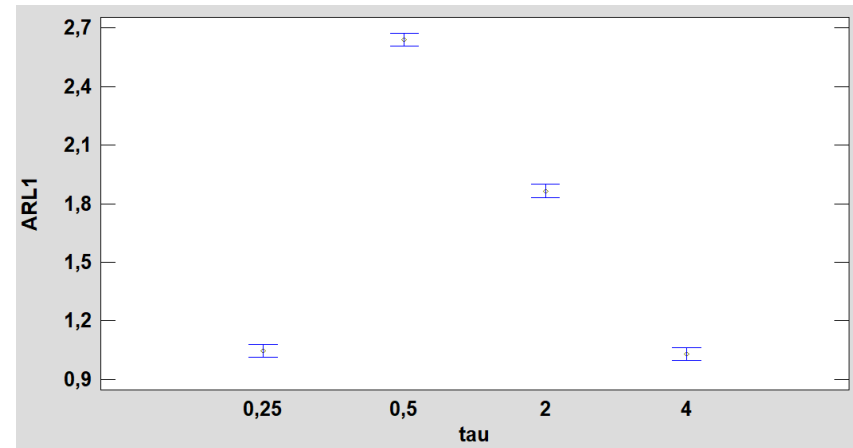
Los resultados numéricos de esta sección se pueden consultar en el anexo B, en el mismo documento Excel donde aparecen los resultados de las secciones 4.1 y 4.2. En este caso, los resultados numéricos se encuentran en la cuarta hoja, llamada "Situación 3".

Para facilitar la interpretación de los resultados, se ha realizado un análisis de varianza (ANOVA) que nos permite identificar como afecta cada uno de los cuatro factores al ARL_1 al igual que ya se ha hecho en las secciones 4.1 y 4.2. La figura 4.3 nos muestra los gráficos de medias obtenidos al realizar el análisis de varianza (ANOVA) junto con unos intervalos LSD al 95%.

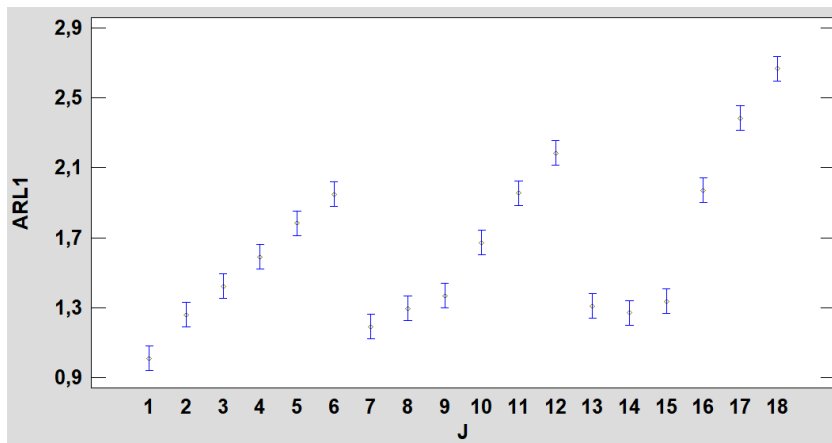
CAPÍTULO 4. RESULTADOS



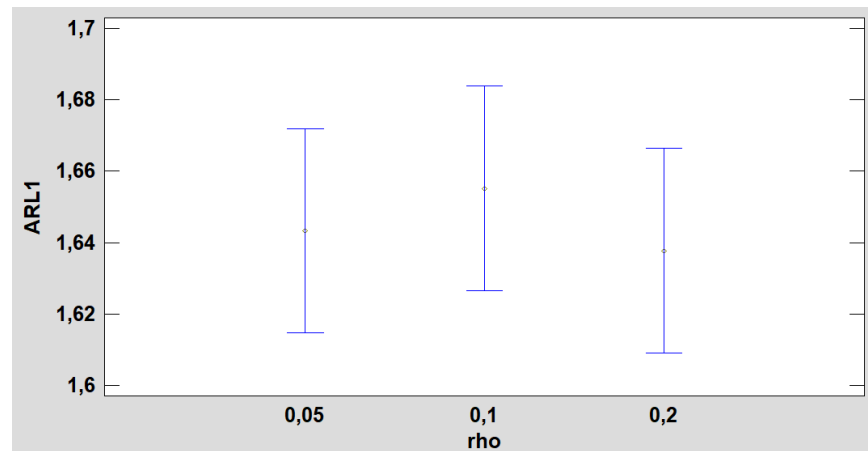
(a) Efectos del tamaño de muestra, n , sobre el ARL_1



(b) Efectos del cambio de dispersión, τ , sobre el ARL_1



(c) Efectos de la distribución de Johnson, J , sobre el ARL_1



(d) Efectos del error de redondeo, ρ , sobre el ARL_1

Figura 4.3 Gráficos de medias e intervalos LSD al 95% para los efectos del ANOVA en la situación 3

La tabla del ANOVA se puede consultar en el anexo C (al igual que sucedía con el ANOVA en las dos secciones anteriores, es decir, las secciones 4.1 y 4.2), la cual toma el nombre de tabla C.3. En esta tabla hallamos que todos los efectos principales son significativos menos el efecto del error de redondeo ρ (ya se observa en la gráfica (d) de la figura 4.3 como los intervalos LSD se encuentran altamente solapados). Respecto a las interacciones de orden 2 también resultan todas significativas menos aquellas interacciones en las que se encuentra involucrado el factor ρ .

En la figura 4.3 se tiene el mismo comportamiento para los factores n, τ y J que ya se tenía en las dos secciones anteriores (secciones 4.1 y 4.2). El comportamiento para estos factores es: para el factor n , tamaño de muestra, vemos que, a mayor tamaño de muestra, menor ARL_1 ; para el factor τ , cambio en la dispersión, se observa que para valores de τ que más disten de 1, menor es el ARL_1 ; para el factor J , distribución de Johnson, se obtiene que, para valores iguales de asimetría, cuanto mayor es la curtosis mayor es el ARL_1 . Respecto al factor ρ , error de redondeo, nos encontramos con que todos los intervalos LSD de la gráfica (d) de la figura 4.3 se encuentran muy solapados por lo que podemos afirmar que este factor no tiene ninguna significancia.

Para finalizar con la sección cabe mencionar que en el anexo D se halla una tabla de medias, la tabla D.1, que contiene el valor promedio de ARL_0 y el valor promedio de ARL_1 para cada nivel y factor.

4.4. Análisis de la situación 2 frente a la situación 1

En esta sección se busca analizar cómo ha influido la presencia de los errores de redondeo en el primer gráfico de control diseñado, el cual no está preparado para la aparición de empates en el instrumento de medida.

Para ello se va a tratar de determinar si en la transición de la situación 1, sin empates, a la situación 2, con empates, el gráfico ha visto incrementado o disminuido tanto el ARL_0 como el ARL_1 . Con el fin de visualizar esto, se construyen tablas que siguen el esquema que nos proporciona la tabla 4.1, donde el valor de A corresponde al número de pruebas del banco de pruebas que experimenta una mejora de ambos ARL , el valor de B1 corresponde al número de pruebas que experimenta una mejora del ARL_0 pero un empeoramiento del ARL_1 , el valor de B2 corresponde al número de pruebas que experimenta un empeoramiento del ARL_0 pero una mejora del ARL_1 , y finalmente el valor C corresponde al número de pruebas que experimenta un empeoramiento para ambos ARL .

CAPÍTULO 4. RESULTADOS

		ARL_1	
		Mejora	Empeora
ARL_0	Mejora	A	B1
	Empeora	B2	C

Tabla 4.1. Esquema que muestra la notación de las secciones 4.4, 4.5 y 4.6 para comparar las tres situaciones entre sí.

Mediante la notación que nos introduce la tabla 4.1, se ha construido la tabla 4.2 que nos informa de como se ha adaptado el gráfico de control diseñado para un escenario ausente de empates (véase sección 3.1) ante un escenario en el que sí hay presencia de empates.

Transiciones (1 → 2)		
A	195	18,06%
B1	156	14,44%
B2	720	66,67%
C	9	0,83%
Total	1080	

Tabla 4.2 Comparativa de los ARL entre la situación 2 y la situación 1.

En la tabla 4.2, se detalla que, en la mayoría de las situaciones, el ARL_0 o el ARL_1 empeora. Justamente de aquí surge la necesidad de diseñar un gráfico de control que tenga implementado los errores de redondeo, evitando así que ambos ARL empeoren cuando hay presencia de empates.

Concretamente, se tienen las siguientes categorías para cada factor:

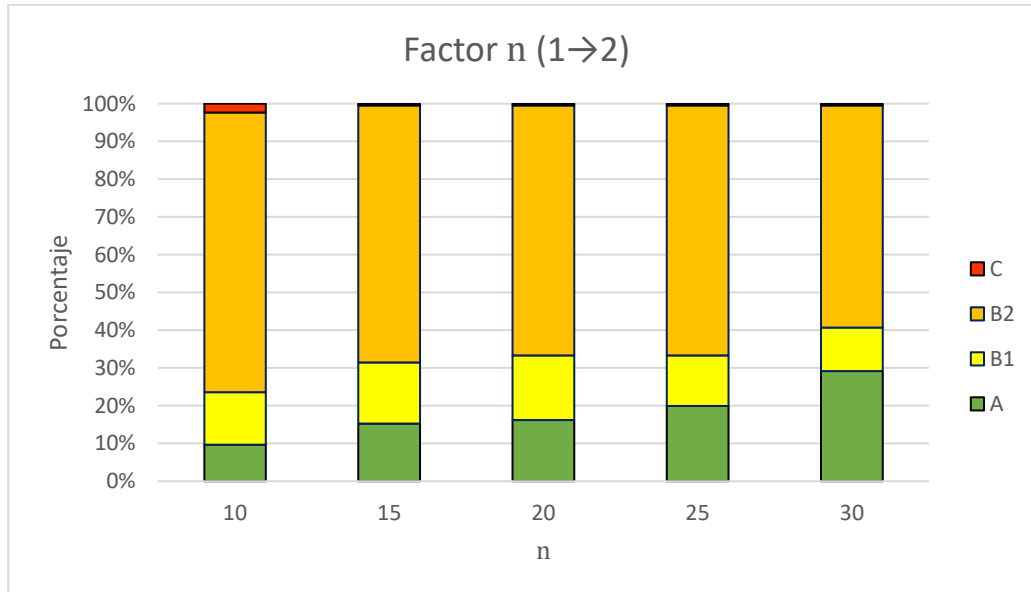


Figura 4.4. Gráfico de barras que compara el factor n en las situaciones 1 y 2.

CAPÍTULO 4. RESULTADOS

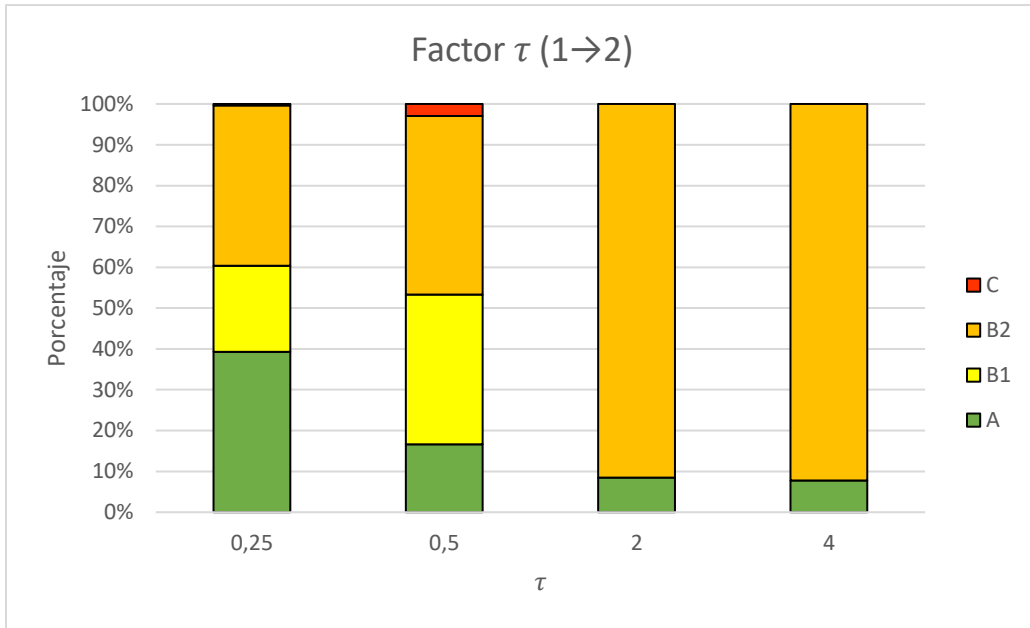


Figura 4.5. Gráfico de barras que compara el factor τ en las situaciones 1 y 2.

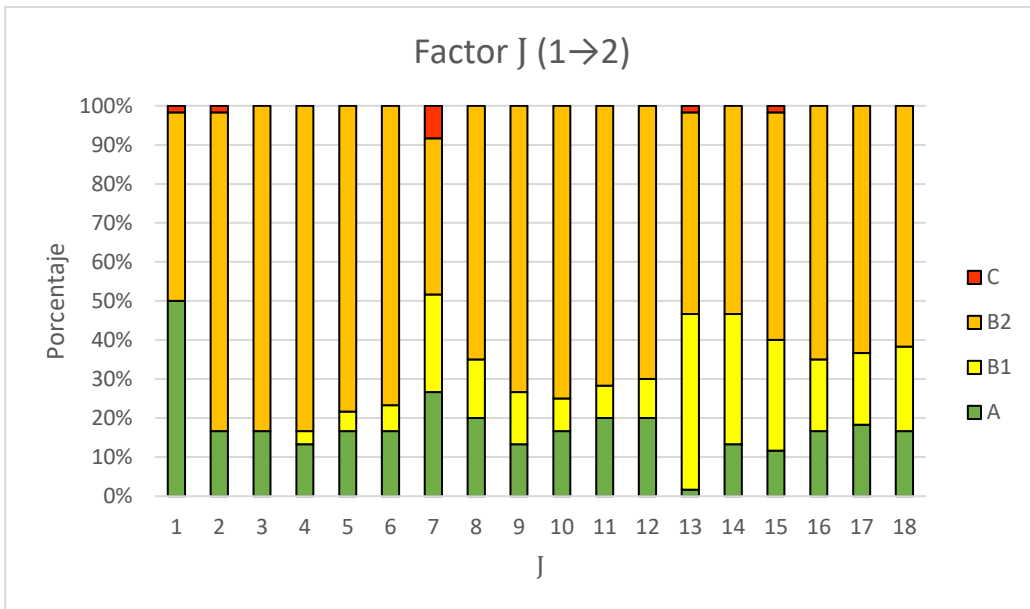


Figura 4.6. Gráfico de barras que compara el factor J en las situaciones 1 y 2.

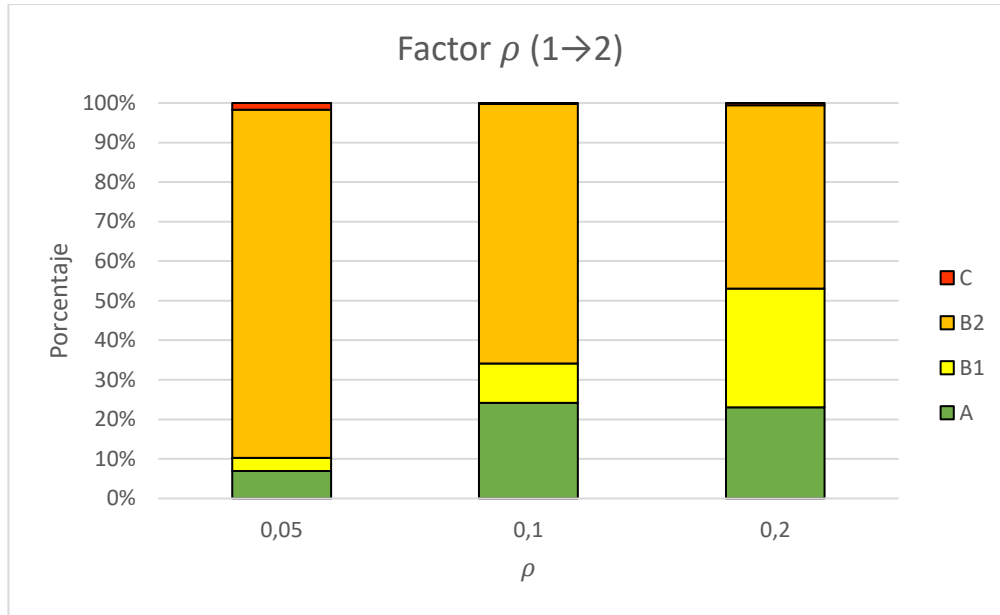


Figura 4.7. Gráfico de barras que compara el factor ρ en las situaciones 1 y 2.

4.5. Análisis de la situación 3 frente a la situación 2

Al comparar la situación 3, gráfico de control diseñado para tratar correctamente la presencia de empates, frente a la situación 2, gráfico de control diseñado para interaccionar con escenarios ausentes de empates pero que se expone a la presencia de empates en el instrumento de medida, se obtiene la tabla 4.3 que sigue la notación mostrada en la tabla 4.1.

Transiciones (2 → 3)		
A	244	22,59%
B1	290	26,85%
B2	542	50,19%
C	4	0,37%
Total	1080	

Tabla 4.3. Comparativa de los ARL entre la situación 3 y la situación 2.

En la tabla 4.3 aparece que en la mayoría de los casos sigue habiendo un empeoramiento bien del ARL_0 o bien del ARL_1 , por lo que se hace necesario analizar las tres situaciones en conjunto para averiguar si realmente está siendo de utilidad el gráfico que trata matemáticamente los empates (véase sección 3.3) lo cual se realiza en la siguiente sección, la sección 4.6.

Para cada factor se tienen las siguientes categorías:

CAPÍTULO 4. RESULTADOS

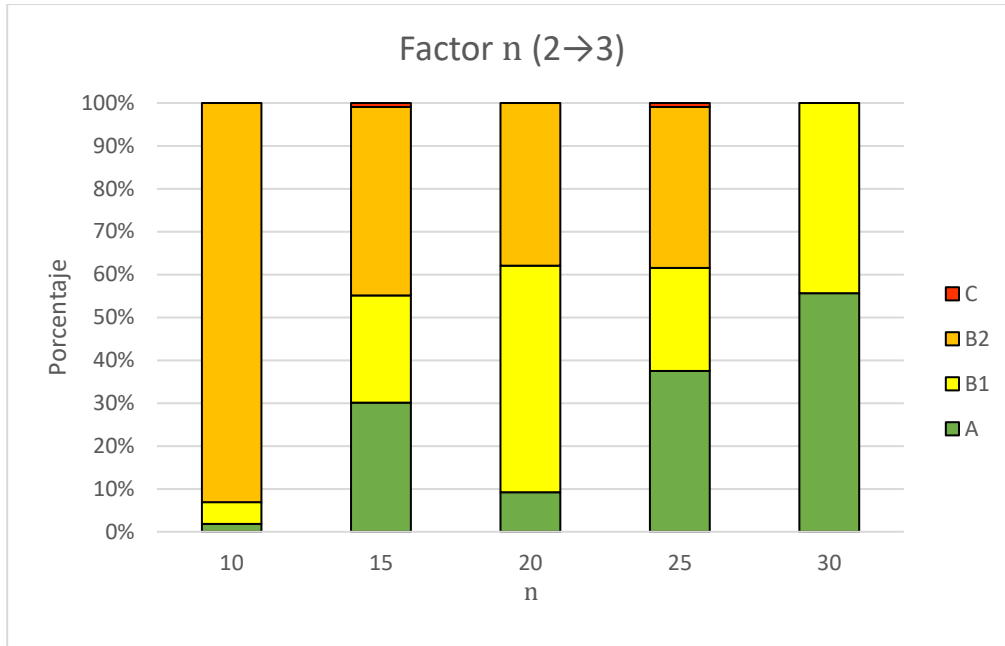


Figura 4.8. Gráfico de barras que compara el factor n en las situaciones 2 y 3.

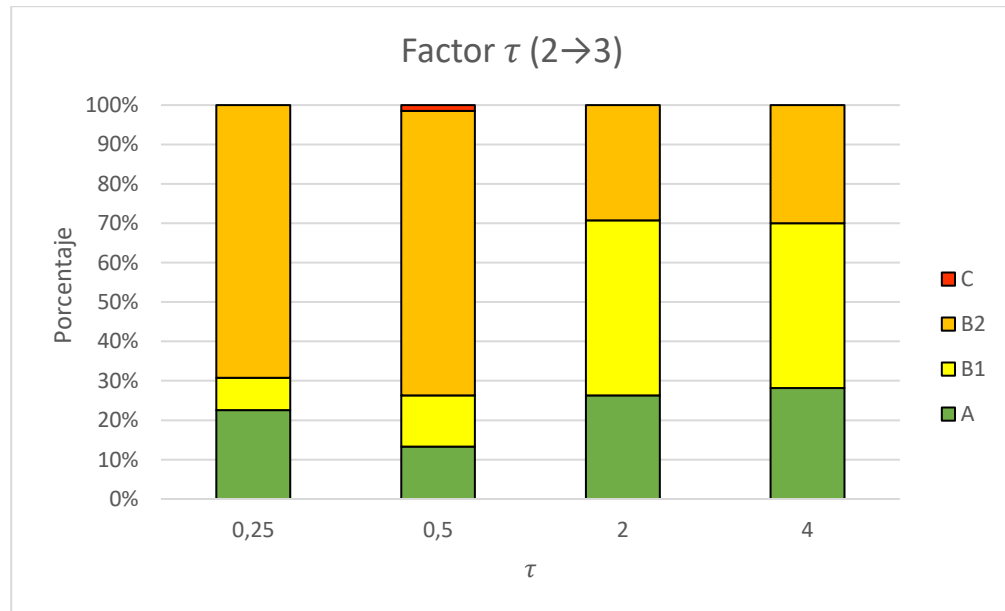


Figura 4.9. Gráfico de barras que compara el factor τ en las situaciones 2 y 3.

CAPÍTULO 4. RESULTADOS

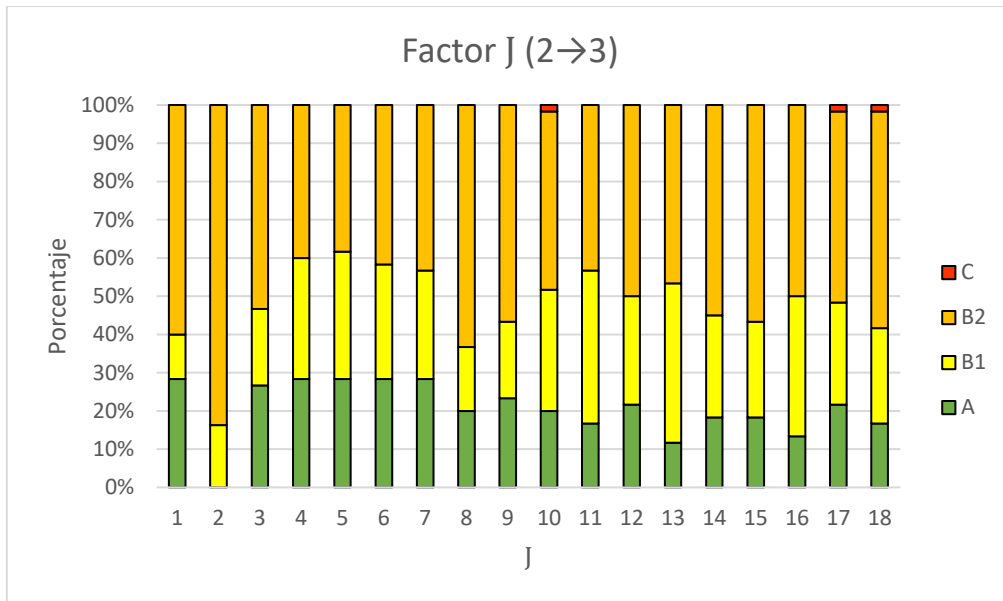


Figura 4.10. Gráfico de barras que compara el factor J en las situaciones 2 y 3.

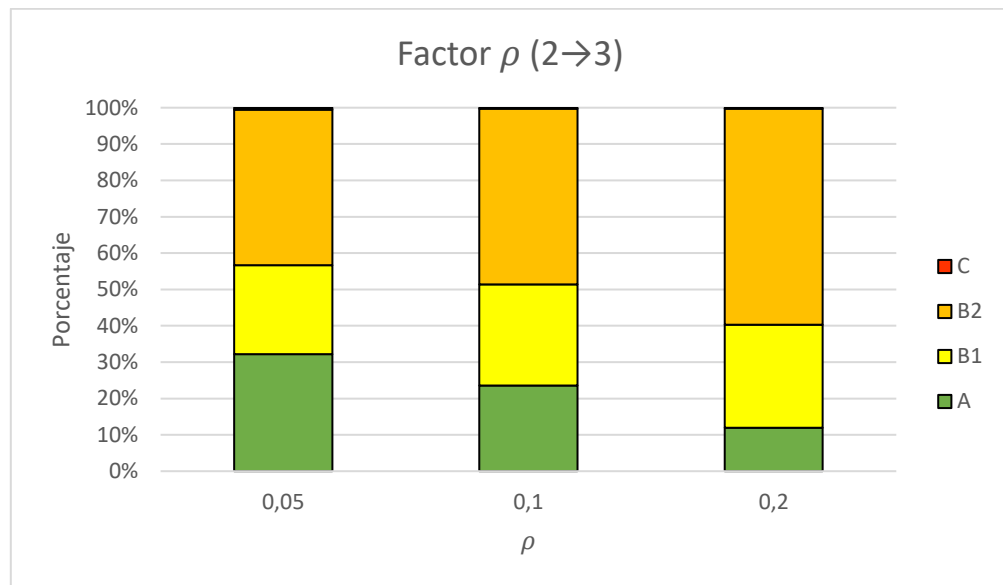


Figura 4.11. Gráfico de barras que compara el factor ρ en las situaciones 2 y 3.

4.6. Análisis comparativo conjunto de las tres situaciones

Para observar si realmente es útil el gráfico de control que procesa los empates, se ha construido la tabla 4.4 donde se observa si para cada una de las pruebas del banco de pruebas se tienen transiciones de la situación 1 a la 2 y de la situación 2 a la 3 que se han obtenido al producirse una mejora o un empeoramiento de uno o ambos de los ARL en cada una de las transiciones.

		Transiciones (2 → 3)			
		A	B1	B2	C
Transiciones (1 → 2)	A	28	23	143	1
	B1	10	0	146	0
	B2	205	266	246	3
	C	1	1	7	0

Tabla 4.4 Comparativa de los ARL entre las tres situaciones.

De la tabla 4.4 se concluye que el gráfico de control diseñado para la presencia de empates o errores de redondeo en el instrumento de medida era completamente necesario pues en la mayoría de casos ubicados como B2 en la transición entre la situación 1 y la situación 2, lo que indica un empeoramiento del ARL_0 , pasan a ser ubicados como caso A o caso B1 por lo que se consigue recuperar ARL_0 perdido e incluso en algunos casos se consigue mejorar no solo el ARL_0 sino también el ARL_1 . Por otra parte, la mayoría de los casos catalogados como B1, que habían perdido ARL_1 , se recuperan al ubicarse posteriormente en B2, por lo que se recuperaría ARL_1 . Finalmente, de los valores catalogados como C, todos abandonan esta categoría para situarse en las otras tres categorías.

Debe tenerse en cuenta que en la situación 3 (es decir, después de *reoptimizar* el gráfico) el ARL_0 vuelve a cumplir los requisitos especificados por el usuario final. Esto significa que, incluso en los casos en los que el ARL_0 en la situación 3 termina siendo peor que en la situación 1, en ningún momento puede ser menor que 370,4 (o la cantidad especificada por el usuario como *input* del modelo de optimización).

4.7. Análisis de la situación 3 frente a la situación 1

En esta sección, la última del capítulo 4, se busca comparar la situación 3, gráfico de control para escenarios con empates en el instrumento de medida, frente a la situación 1, gráfico de control para escenarios sin empates en el instrumento de medida.

Para ello se realiza un análisis de varianza (ANOVA) a diferencia de lo realizado en las comparaciones anteriores, llevadas a cabo en las secciones 4.4, 4.5 y 4.6. De este modo los ANOVAS se realizan con los siguientes cuatro factores: n , tamaño de muestra; τ , cambio en la dispersión; J , distribución de Johnson subyacente; y ρ , resolución del instrumento de medida lo que determina el error de redondeo.

Por otro lado, se proponen dos variables dependientes para realizar el ANOVA multifactorial, $\Delta ARL_1^{(3-1)}$ y $\% \Delta ARL_1^{(3-1)}$, que responden a las ecuaciones (4.1) y (4.2) mostradas a continuación.

$$\Delta ARL_1^{(3-1)} = ARL_1^{(3)} - ARL_1^{(1)} \quad (4.1)$$

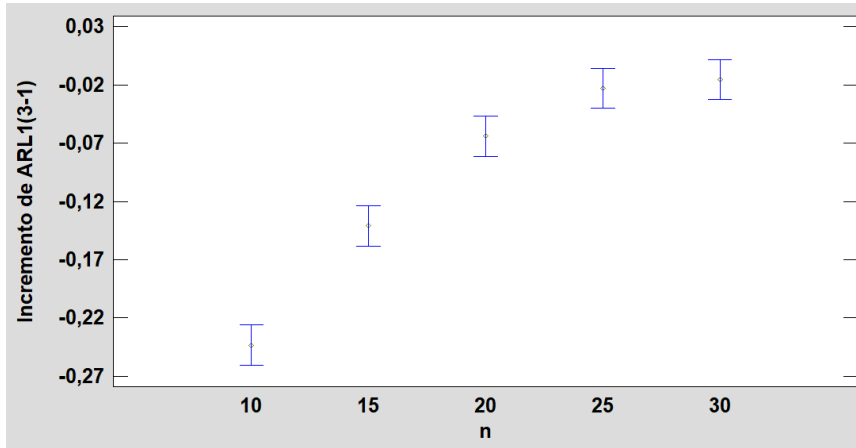
$$\% \Delta ARL_1^{(3-1)} = \frac{ARL_1^{(3)} - ARL_1^{(1)}}{ARL_1^{(1)}} \quad (4.2)$$

En las ecuaciones (4.1) y (4.2), el superíndice para cada ARL_1 sirve para referirse a la situación a la que corresponde ese ARL_1 . Así, $ARL_1^{(1)}$ corresponde al ARL_1 en la situación 1 y $ARL_1^{(3)}$ corresponde al ARL_1 en la situación 3.

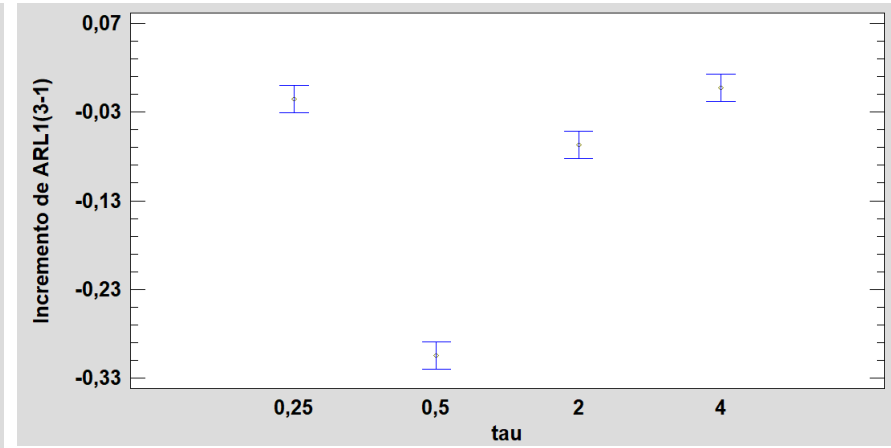
Aunque utilizando tanto $\Delta ARL_1^{(3-1)}$ como $\% \Delta ARL_1^{(3-1)}$ es posible realizar el análisis de varianza (de hecho, se ha realizado con ambas expresiones), en este trabajo se decide mostrar únicamente los gráficos de medias para $\Delta ARL_1^{(3-1)}$ en la figura 4.12. Además, la tabla del ANOVA, tabla C.4, se muestra en el anexo C y la tabla de medias, tabla D.4, aparece en el anexo D.

A partir del ANOVA se puede determinar cómo afecta cada uno de los cuatro factores al gráfico de control diseñado para la presencia de empates al observar la presencia de estos y determinando así si conserva la misma eficacia que se observaba en el gráfico de control diseñado para la ausencia de empates en ausencia de estos.

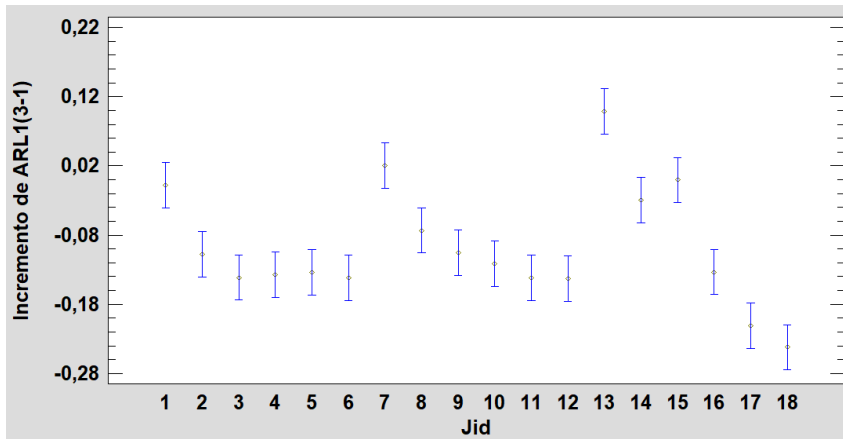
CAPÍTULO 4. RESULTADOS



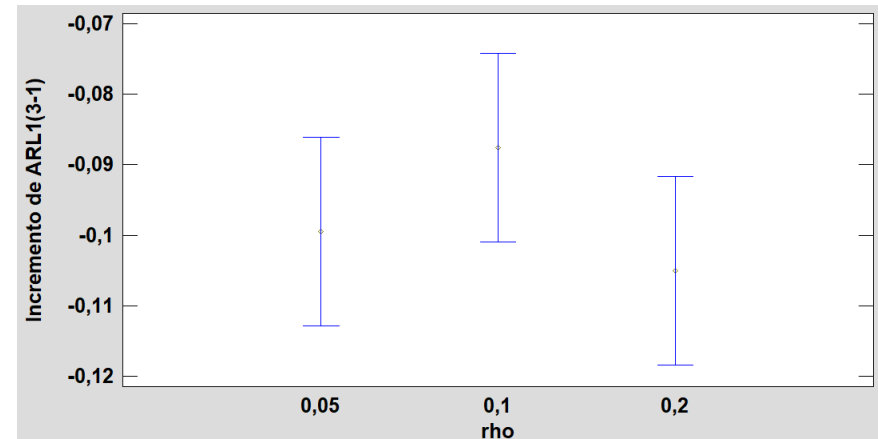
(a) Efectos del tamaño de muestra, n , sobre el $\Delta ARL_1^{(3-1)}$



(b) Efectos del cambio de dispersión, τ , sobre el $\Delta ARL_1^{(3-1)}$



(c) Efectos de la distribución, J , sobre el $\Delta ARL_1^{(3-1)}$



(d) Efectos del error de redondeo, ρ , sobre el $\Delta ARL_1^{(3-1)}$

Figura 4.12. Gráficos de medias e intervalos LSD al 95% para los efectos del ANOVA en la situación 3 vs situación 1.

En la figura 4.12 se observa que el gráfico sufre una muy ligera pero despreciable pérdida de eficiencia al implementar la presencia de empates y tener que enfrentarse a los mismos.

Por otro lado, hay que mencionar que todos los efectos principales son significativos excepto el efecto del factor ρ , resolución del instrumento de medida, y que todas las interacciones de segundo orden son significativas excepto aquellas en las que interviene el factor ρ .

Así, analicemos que se extrae de la figura 4.12 para cada factor: para el factor n se observa que para un mayor tamaño de muestra mayor es $\Delta ARL_1^{(3-1)}$ sin llegar a sobrepasar $\Delta ARL_1^{(3-1)} = 0$ por lo que para todos los niveles $ARL_1^{(1)}$ es mayor que $ARL_1^{(3)}$; para el factor τ se extrae que para cambios en la dispersión pequeños el gráfico de la situación 3 posee menor ARL_1 que en la situación 1 mientras que para cambios grandes el desempeño es el mismo; para el factor J se observa que para distribuciones con una asimetría idéntica cuanto menor es la curtosis menor es la diferencia entre la situaciones 3 y 1; y para el factor ρ se tiene que los intervalos LSD se encuentran altamente solapados por lo que no hay diferencias significativas entre ellas situándose sus medias en un valor ligeramente inferior a 0.

Con esto termina el capítulo 4, donde se ha conseguido analizar las tres situaciones por separado y también se ha conseguido compararlas entre ellas.

5. Conclusiones

En este último capítulo se busca analizar si los objetivos propuestos en la sección 1.4 se han alcanzado. Por lo tanto, se procede a analizar cada uno de los objetivos por separado para analizar si se han cumplido con éxito.

1. Primeramente, se ha conseguido estudiar el gráfico de control no paramétrico para la dispersión basado en el estadístico de signo en ausencia de empates de Pawar et al. (2018), diseñándose su implementación mediante el software Scilab (véase anexo A) y se han determinado sus parámetros óptimos para un banco de pruebas. Posteriormente, se ha analizado la influencia de las condiciones iniciales en el comportamiento del gráfico (véase sección 4.1).
2. En segundo lugar, se ha analizado el impacto de la existencia de errores de redondeo en el instrumento de medida, observándose una degradación del gráfico no paramétrico para la dispersión objeto de estudio, pues no está diseñado para la presencia de empates. En concreto, sobre el banco de pruebas se ha observado que, en un 81,94% de los casos estudiados, los valores reales del ARL_1 o del ARL_0 , o de ambos, eran peores de lo esperado. Incluso se observa que, en algunos casos, dependiendo de la forma de la distribución de la característica de calidad, objeto de estudio, los resultados obtenidos en términos de ARL_0 y ARL_1 pueden ser muy extremos (como es el caso de la distribución $J = 13$ del banco de pruebas considerado).
3. Por último, se ha diseñado un procedimiento para la determinación óptima de los parámetros del gráfico objeto de estudio asumiendo la presencia de errores de redondeo, el cual funciona con una eficiencia bastante similar al gráfico primero, pero ya adaptado para procesar escenarios con presencia de empates. En concreto, sobre el banco de pruebas analizado, se ha observado que en los gráficos *reoptimizados* (es decir, teniendo en cuenta la probabilidad de empate) se consiguen valores de ARL_1 muy similares a los que teóricamente se podrían haber alcanzado en ausencia de empates, todo ello respetando el nivel de falsas alarmas deseado.

Además de cumplirse los objetivos es importante destacar que, por lo general, los gráficos de control diseñados funcionan con mejor eficiencia (es decir, menor ARL_1) si se utilizan con un tamaño de muestra elevado, un cambio de dispersión elevado, una distribución subyacente con una curtosis baja. El error de redondeo no resulta ser significativo por lo que respecta a su influencia con los ARL (eso sí, es importante utilizar el gráfico de control diseñado para la presencia de empates siempre y cuando aparezcan, lo cual resulta lo más realista).

En concreto, las pruebas realizadas utilizando distintas distribuciones para la característica de calidad objeto de estudio, con diferentes valores de asimetría y curtosis, permiten concluir que, en el caso de asumir erróneamente un

comportamiento gaussiano para dicha característica, el valor de ARL_1 que se obtendría sería, en realidad, mayor del esperado si la distribución posee mayor curtosis que la normal (o sea, es leptocúrtica), o menor si posee menor curtosis (platicúrtica). En cambio, no se han observado diferencias en este sentido asociadas a la mayor o menor asimetría. En definitiva, las experiencias llevadas a cabo evidencian la necesidad de conocer algunas características de la distribución del proceso, más allá de su media y su desviación típica, para saber realmente cuál será el comportamiento del gráfico optimizado (con o sin presencia de empates).

Ya para concluir, hay que señalar que una posible mejora de los algoritmos tratados sería buscar la forma de conseguir una optimización eficiente sin que p_0 se encuentre discretizado, pudiendo así obtener una optimalidad real y completa.

Bibliografía

Alfaro-Cortés E, Alfaro Navarro J, Mondéjar-Jiménez J, Vargas-Vargas M (2004). *Control estadístico de la calidad: una breve reseña histórica*. Universidad de Castilla-La Mancha.

Alloway JA, Raghavachari M (1991). Control Chart Based on the Hodges-Lehmann Estimator. *Journal of Quality Technology*, 23:336–347.

Amin RW, Reynolds MR, Saad B (1995). Nonparametric quality control charts based on the sign statistic. *Communications in Statistics - Theory and Methods*, 24(6):1597–1623.

Balakrishnan N, Lai CD (2009). *Continuous Bivariate Distributions*. 2.^aed. Springer, Dordrecht.

Bradley JV (1968). *Distribution-free statistical tests*. 1.^aed. Prentice-Hall, New Jersey.

Brooks C, Clare A, Dalle Molle J, Persaud G (2005). A comparison of extreme value theory approaches for determining value at risk. *Journal of Empirical Finance*, 12(2):339–352.

Castagliola P, Celano G, Fichera S, Guiffrida F (2007). A variable sampling Interval S2-EWMA control chart for monitoring the process variance. *International Journal of Technology Management*, 37(1-2):125-146.

Castagliola P, Tran KP, Celano G, Maravelakis PE (2020). The Shewhart Sign Chart with Ties: Performance and Alternatives. En: Koutras M, Triantafyllou I (eds.), *Distribution-Free Methods for Statistical Process Monitoring and Control*. Springer, Cham.

Castagliola P, Tran KP, Celano G, Rakitzis AC, Maravelakis PE (2019). An EWMA-type sign chart with exact run length properties. *Journal of Quality Technology*, 51(1):51–63.

Chakraborti S, Van der Laan P, Bakir ST (2001). Nonparametric control charts: An overview and some results. *Journal of Quality Technology*, 33(3):304–315.

Collani EV, Sheil J (1989). An approach to controlling process variability. *Journal of Quality Technology*, 21:87-96.

Crowder SV, Hamilton MD (1992). An EWMA for monitoring a process standard deviation. *Journal of Quality Technology*, 24(1):12-21.

Ferrer-Riquelme AJ (2009). Statistical Control of Measures and Processes. En: Brown SD, Tauler Ferré R, Walczak B (eds.), *Comprehensive Chemometrics. Chemical and Biochemical Data Analysis*, cap. 1.04, págs. 97–126. Elsevier, Amsterdam.

Flynn MR (2006). Fitting human exposure data with the Johnson SB distribution. *Journal of Exposure Science and Environmental Epidemiology*, 16(1):56–62.

Gibbons JD, Chakraborti S (2003). *Nonparametric Statistical Inference*, 4.^aed. Dekker, New York.

Hawkins DM (1981). A Cusum for a scale parameter. *Journal of Quality Technology*, 13(4):228-231.

He D, Grigoryan A (2002). Construction of double sampling s-control chart for agile manufacturing. *Quality and Reliability Engineering International*. 18:343-355.

Huang HJ, Chen FL (2005). A synthetic control chart for monitoring process dispersion with sample standard deviation. *Computers & Industrial Engineering*, 49(2):221-240.

Johnson N (1949). Bivariate Distributions Based on Simple Translation Systems. *Biometrika*, 36(3-4):297-304.

Juran JM, Godfrey AB (2001). *Manual de calidad de Juran*. McGraw-Hill Education, New York.

Kang CW, Lee MS, Seon YJ, Hawkins DM (2007). A control chart for the coefficient of variation. *Journal of Quality Technology*, 39(2):151-158.

Langenberg P, Iglewicz B (1986). Trimmed Mean X and R Charts. *Journal of Quality Technology*, 18:152–161.

Linna KW, Woodall WH (2001). Effect of measurement error on Shewhart control charts. *Journal of Quality Technology*, 33(2):213–222.

Lowry CA, Champ CW, Woodall WH (1995). The performance of control charts for monitoring process variation. *Communications in Statistics – Simulation and Computation*, 24:409-437.

Maleki MR, Amiri A, Castagliola P (2017). Measurement errors in statistical process monitoring: A literature review. *Computers & Industrial Engineering*, 103:316–329.

Montgomery DC (2012). *Introduction to Statistical Quality Control*. 7.^aed. John Wiley & Sons, Hoboken, NJ.

Nelson LS (1990). Monitoring reduction in variation with a range chart. *Journal of Quality Technology*, 22:163-165.

Page ES (1963). Controlling the Standard Deviation by Cusums and Warning Lines. *Technometrics*, 5:307-315.

Pawar VY, Shirke, DT, Khilare SK (2018). A Nonparametric Control Chart for Process Variability Based on Quantiles. *International Journal of Statistics and Economics*, 19(3).

Riaz M (2008). A Dispersion Control Chart. *Communications in Statistics – Simulation and Computation*, 37(6):1239-1261.

Shewhart WA (1931). *Economic Control of Quality of Manufactured Product*. Van Nostrand, New York.

Shewhart WA, Deming WE (1986). *Statistical method from the viewpoint of quality control*. Dover, New York.

Yeong W, Khoo MBC, Tham L, Teoh W, Rahim M (2017). Monitoring the coefficient of variation using a variable sampling interval EWMA chart. *Journal of Quality Technology*, 49(4):380-401.

Yourstone SA, Zimmer WJ (1992). Nonnormality and the Design of Control Charts for Averages. *Decision Sciences*, 23:1099–1113.

Zorriassatine F, Tannock JDT (1998). A review of neural networks for statistical process control. *Journal of Intelligent Manufacturing*, 9(3):209–224.

Anexos

Anexo A. Código de Scilab

El código de Scilab que se ha utilizado para obtener todos los resultados numéricos de este trabajo final de máster se encuentra disponible en formato. sce en el siguiente enlace: <https://bit.ly/3qNnJFQ>

Para ejecutar el código con éxito se han utilizado algunas funciones ya programadas que se encuentran en la siguiente carpeta: <https://bit.ly/3xgcYhD>

Anexo B. Archivo Excel con los resultados numéricos

Los resultados numéricos obtenidos al ejecutar los diferentes algoritmos con el banco de pruebas se encuentran en un archivo Excel al que se puede acceder mediante el siguiente enlace: <https://bit.ly/3qJyl8m>

Anexo C. Tablas de los ANOVAS de las tres situaciones y de la comparación de la situación 3 frente a la situación 1

Fuente	SC	Gl	CM	Razón-F	Valor-P
Efectos principales					
n	175,67	4	43,9174	162,21	0
τ	220,464	3	73,4881	271,43	0
J	94,6434	17	5,56726	20,56	0
Interacciones					
n x τ	191,568	12	15,964	58,96	0
n x J	59,494	68	0,874912	3,23	0
τ x J	90,7679	51	1,77976	6,57	0
Residuos	55,2317	204	0,270744		
Total (Corr.)	887,839	359			

Tabla C.1. Tabla del ANOVA de los tres factores n, τ y J sobre el ARL_1 para la situación 1.

Fuente	SC	Gl	CM	Razón-F	Valor-P
Efectos principales					
n	679,386	4	169,846	184,86	0
τ	987,664	3	329,221	358,33	0
J	321,267	16	20,0792	21,85	0
ρ	54,3009	2	27,1504	29,55	0
Interacciones					
n x τ	893,447	12	74,4539	81,04	0
n x J	217,978	64	3,40591	3,71	0
n x ρ	57,1968	8	7,1496	7,78	0
τ x J	355,728	48	7,41099	8,07	0
τ x ρ	152,256	6	25,3759	27,62	0
J x ρ	63,0117	32	1,96912	2,14	0,0003
Residuos	757,061	824	0,918764		
Total (Corr.)	4539,3	1019			

Tabla C.2. Tabla del ANOVA de los cuatro factores n, τ , J y ρ sobre el ARL_1 para la situación 2.

ANEXOS

Fuente	SC	Gl	CM	Razón-F	Valor-P
Efectos principales					
n	408,063	4	102,016	662,12	0
τ	479,485	3	159,828	1037,35	0
J	213,725	17	12,5721	81,6	0
ρ	0,0571522	2	0,0285761	0,19	0,8307
Interacciones					
n x τ	457,16	12	38,0967	247,26	0
n x J	129,918	68	1,91056	12,4	0
n x ρ	0,167225	8	0,0209031	0,14	0,9976
τ x J	197,289	51	3,86842	25,11	0
τ x ρ	0,0786774	6	0,0131129	0,09	0,9977
J x ρ	1,33473	34	0,0392567	0,25	1
Residuos	134,66	874	0,154074		
Total (Corr.)	2021,94	1079			

Tabla C.3. Tabla del ANOVA de los cuatro factores n, τ , J y ρ sobre el ARL_1 para la situación 3.

Fuente	SC	Gl	CM	Razón-F	Valor-P
Efectos principales					
n	7,9076	4	1,9769	58,82	0
τ	16,1015	3	5,36718	159,69	0
J	7,31455	17	0,430268	12,8	0
ρ	0,0571522	2	0,0285761	0,85	0,4277
Interacciones					
n x τ	10,8927	12	0,907725	27,01	0
n x J	7,22374	68	0,106231	3,16	0
n x ρ	0,167225	8	0,0209031	0,62	0,7599
τ x J	11,3635	51	0,222814	6,63	0
τ x ρ	0,0786772	6	0,0131129	0,39	0,8856
J x ρ	1,33473	34	0,0392567	1,17	0,2361
Residuos	29,3744	874	0,0336091		
Total (Corr.)	91,8158	1079			

Tabla C.4. Tabla del ANOVA de los cuatro factores n, τ , J y ρ sobre el ΔARL_1 para la situación 3 vs situación 1.

Anexo D. Tablas de medias para cada una de las tres situaciones y para la comparación de la situación 3 frente a la situación 1

Factor	Nivel	ARL ₀ promedio	ARL ₁ promedio
n	10	744,236	3,04903
	15	815,905	1,85571
	20	558,054	1,44273
	25	875,647	1,22943
	30	760,125	1,13676
τ	0,25	793,345	1,06152
	0,5	695,066	2,9449
	2	729,632	1,93206
	4	785,131	1,03244
J	1	1822,44	1,0169
	2	1464,7	1,36564
	3	708,184	1,56348
	4	548,378	1,72862
	5	523,016	1,91816
	6	504,984	2,08971
	7	890,23	1,16941
	8	782,802	1,36755
	9	708,9	1,47363
	10	521,487	1,79353
	11	504,984	2,09729
	12	500,447	2,32803
	13	939,019	1,20896
	14	800,733	1,30013
15	762,974	1,33746	
16	507,358	2,1042	
17	504,451	2,59671	
18	519,192	2,90974	

Tabla D.1. Tabla de medias para cada factor y cada nivel para la situación 1.

ANEXOS

Factor	Nivel	ARL ₀ promedio	ARL ₁ promedio
n	10	828,098	3,37881
	15	1457,95	2,00258
	20	810,402	1,52482
	25	1128,83	1,24738
	30	1036,8	1,14198
τ	0,25	1865,05	1,09889
	0,5	956,223	3,47291
	2	667,284	1,83646
	4	721,103	1,02821
J	1	2786,81	1,01777
	2	1356,82	1,30788
	3	634,785	1,49711
	4	515,884	1,6522
	5	510,869	1,83537
	6	512,12	2,00668
	7	1426,9	1,41378
	8	812,345	1,37874
	9	706,857	1,45201
	10	534,301	1,74918
	11	542,886	2,04721
	12	575,058	2,30207
	14	2920,13	1,86469
	15	1000,85	1,63465
16	804,155	2,29304	
17	931,586	2,84572	
18	1318,69	3,30685	
ρ	0,05	656,712	1,67654
	0,1	719,442	1,71619
	0,2	1781,09	2,18461

Tabla D.2. Tabla de medias para cada factor y cada nivel para la situación 2.

ANEXOS

Factor	Nivel	ARL ₀ promedio	ARL ₁ promedio
n	10	628,295	2,80557
	15	658,99	1,7148
	20	670,616	1,37877
	25	632,407	1,20646
	30	766,977	1,12136
τ	0,25	861,893	1,04641
	0,5	584,958	2,6402
	2	602,121	1,86527
	4	636,855	1,02968
J	1	1292,15	1,00913
	2	1393,31	1,25785
	3	591,748	1,42221
	4	553,391	1,59094
	5	530,151	1,7839
	6	496,396	1,94804
	7	1115,31	1,18985
	8	617,47	1,29432
	9	626,079	1,36847
	10	515,654	1,67207
	11	502,069	1,95576
	12	506,857	2,18484
	13	598,336	1,30786
	14	603,627	1,27079
	15	601,851	1,33702
16	529,678	1,97085	
17	501,567	2,38561	
18	510,575	2,66756	
ρ	0,05	637,117	1,6433
	0,1	713,941	1,65516
	0,2	663,313	1,63771

Tabla D.3. Tabla de medias para cada factor y cada nivel para la situación 3.

Factor	Nivel	ΔARL_1 promedio
n	10	-0,24346
	15	-0,140907
	20	-0,0639587
	25	-0,0229689
	30	-0,0153974
τ	0,25	-0,0151135
	0,5	-0,304692
	2	-0,0667929
	4	-0,0027559
J	1	-0,00777631
	2	-0,107794
	3	-0,14127
	4	-0,137681
	5	-0,134257
	6	-0,141675
	7	0,0204437
	8	-0,0732232
	9	-0,10516
	10	-0,121453
	11	-0,141534
	12	-0,143196
	13	0,0988948
	14	-0,0293414
	15	-0,000437624
	16	-0,133353
	17	-0,211103
	18	-0,24218
ρ	0,05	-0,0994291
	0,1	-0,0875698
	0,2	-0,105017

Tabla D.4. Tabla de medias para cada factor y cada nivel para la situación 3 vs situación 1.