

*INTERVALOS DE PREDICCIÓN PARA EL
TIEMPO DE VIDA DE PRODUCTOS EN
OPERACIÓN*

ALEJANDRA ESTEFANÍA PATIÑO HOYOS
INGENIERA INDUSTRIAL
CÓDIGO: 201022922



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
ESCUELA DE ESTADÍSTICA
MEDELLÍN
NOVIEMBRE DE 2012

INTERVALOS DE PREDICCIÓN PARA EL TIEMPO DE VIDA DE PRODUCTOS EN OPERACIÓN

ALEJANDRA ESTEFANÍA PATIÑO HOYOS
INGENIERA INDUSTRIAL
CÓDIGO: 201022922

TRABAJO DE TESIS PARA OPTAR AL TÍTULO DE
MAGISTER EN CIENCIAS-ESTADÍSTICA

DIRECTOR
NELFI GERTRUDIS GONZÁLEZ ÁLVAREZ
DOCTORA EN ESTADÍSTICA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
ESCUELA DE ESTADÍSTICA
MEDELLÍN
NOVIEMBRE DE 2012

Título en español

Intervalos de predicción para el tiempo de vida de productos en operación

Title in English

Prediction intervals for the lifetime of products in operation

Resumen: Actualmente los fabricantes sienten una fuerte presión por desarrollar nuevos y mejores productos en plazos más cortos, garantizando su confiabilidad en manos del cliente y asegurando una alta calidad para ser competitivos en el mercado. Muchos de estos requerimientos exigen la realización de pruebas de vida acelerada (ALT) en laboratorio que proporcionen evaluaciones en corto tiempo de la confiabilidad de los materiales, componentes y subsistemas del producto así como para predecir su desempeño en condiciones normales de uso. Sin embargo, resulta más adecuado combinar los resultados de pruebas de laboratorio con los datos de campo mediante la formulación de un modelo apropiado que permita establecer la relación entre ambos tipos de datos y aprovecharlos en la determinación y estimación del modelo de vida actual para luego proyectar o predecir los efectos de cambios en el diseño del producto sobre el desempeño futuro del mismo. En este trabajo se pretenden construir intervalos de predicción para el tiempo de vida de productos en operación a partir del modelo de tiempo de falla propuesto por Meeker *et al.* (2009), el cual es una mezcla de distribuciones donde se combina información de la distribución de los ciclos hasta la falla en una ALT, la cual pertenece a la familia de log-localización-escala (Weibull o lognormal), con un modelo de tasa de uso del producto en el campo. Las técnicas a considerar para la predicción serán las propuestas por Meeker y Escobar (1998) y Escobar y Meeker (1999). Se realizará también una comparación vía simulación de las distintas técnicas de construcción de intervalos de predicción comúnmente utilizadas en problemas de confiabilidad.

Abstract: Currently manufacturers feel intense pressure to develop new and better products in shorter time, ensuring reliability in the hands of the customer and high quality to be competitive in the market. Many of these requirements need the implementation of Accelerated Life Tests (ALT) in the laboratory to provide short-term assessments of the reliability of the materials, components and subsystems of the product and to predict their performance under normal conditions of use. However, it is more appropriate to combine the results of laboratory tests with field data by formulating an appropriate model to establish the relationship between both types of data and use them in the identification and estimation of the model of present life then project or predict the effects of changes in product design for its future performance. In this paper we seek to construct prediction intervals for the lifetime of products in field from the failure time model proposed by Meeker *et al.* (2009), which is a mixture of distributions which combines information of the cycles to failure distribution in an ALT which belongs to the family of log-location-scale (Weibull and lognormal) with a use rate model of product in the field. The techniques to consider in the prediction are the proposed by Meeker and Escobar (1998) and Escobar and Meeker (1999). A comparison is also performed via simulation of different prediction intervals construction techniques commonly used in reliability problems.

Palabras clave: Confiabilidad, Distribuciones de log-localización-escala, Pruebas de Vida Aceleradas, Modelo de Tasa de Uso, Intervalos de Predicción, Garantía.

Keywords: Reliability, Log-location-scale distribution, Accelerated Life Tests, Use Rate Model, Prediction Intervals, Warranty.

Dedicado a

Mi mamá, mi papá y mi hermana.

Agradecimientos

Agradezco a todas las personas que hicieron posible el desarrollo exitoso de esta investigación, entre ellos a la profesora Nelfi González por su valioso trabajo como asesora y a los profesores Luis A. Escobar y Juan Carlos Correa quienes con sus aportes contribuyeron en el cumplimiento de los objetivos de este proyecto.

También agradezco a la Escuela de Estadística de la Facultad de Ciencias de la Universidad Nacional de Colombia Sede Medellín por la colaboración prestada al facilitar los equipos de computo donde se realizaron los estudios de simulación.

Por último agradezco a mis padres por su incondicional apoyo y comprensión así como a Sandra Restrepo por su amistad e importante acompañamiento durante este proceso.

Índice general

Índice general	I
Índice de tablas	IV
Índice de figuras	VI
Introducción	IX
1. Marco Teórico	1
1.1. Pruebas de Vida Acelerada (ALT)	1
1.1.1. Relación entre los resultados de la ALT y la confiabilidad en campo	1
1.2. Intensidades de uso de los productos en manos del cliente y su influencia sobre las políticas de garantía	2
1.2.1. Modelación de la intensidad de uso	2
1.3. Modelos de mezclas finitas	4
1.3.1. Definición básica	4
1.3.2. Interpretación de los modelos de mezcla	5
1.3.3. Estructura de datos incompletos en el problema de mezcla	5
1.3.4. Formulación paramétrica del modelo de mezcla	6
1.3.5. Estimación por máxima verosimilitud para el modelo de mezcla	7
1.4. El modelo de tasa de uso para confiabilidad en campo	8
1.4.1. Estimación por máxima verosimilitud para el modelo de tasa de uso con datos de vida sin censura	11
1.4.2. Estimación por máxima verosimilitud para el modelo de tasa de uso con datos de vida censurados	13
1.5. Predicción estadística	15
1.5.1. Intervalos de predicción	16

1.5.2.	Intervalos de predicción bajo el marco frecuentista	16
1.5.2.1.	Método plug-in o naive	17
1.5.3.	Distribuciones predictivas	17
1.5.4.	Calibración	18
1.6.	Predicción estadística basada en datos de vida censurados	20
1.6.1.	Conceptos sobre los intervalos de predicción	20
1.6.1.1.	Intervalos de predicción para tiempos de falla (dado θ)	20
1.6.1.2.	Probabilidad de cobertura para procedimientos de intervalos de predicción estadística (θ estimado)	21
1.6.1.3.	Relación entre límites de predicción unilaterales e intervalos de predicción bilaterales	22
1.6.1.4.	El método naive (plug-in) para calcular intervalos de predicción estadística	22
1.6.2.	Calibración de límites de predicción estadística naive (plug-in)	23
1.6.2.1.	Calibración aproximada de los límites de predicción estadística naive	23
1.6.2.2.	Método 1: Calibración por simulación del proceso de muestreo/predicción	24
1.6.2.3.	Equivalencia entre el procedimiento de calibración y el procedimiento pivotal (aproximadamente pivotal) para intervalos de predicción a partir de distribuciones de log-localización-escala	29
1.6.2.4.	Método 3: Calibración promediando las probabilidades de cobertura condicional	30
2.	Predicción estadística para el modelo de tasa de uso	33
2.1.	Un problema con el componente A	33
2.1.1.	Simulación de los datos para el problema con el componente A	33
2.1.2.	Estimación por máxima verosimilitud para el modelo de tasa de uso para el componente A	36
2.1.2.1.	Estimación por máxima verosimilitud para el modelo de tasa de uso para el componente A con datos de vida sin censura	37
2.1.2.2.	Estimación por máxima verosimilitud para el modelo de tasa de uso para el componente A con datos de vida censurados	38
2.1.3.	Predicción estadística para el modelo de tasa de uso para el componente A	43

2.1.3.1. Calibración de los límites de predicción unilaterales naive para el modelo de tasa de uso para el componente A	47
3. Predicción estadística para el componente A con información sólo de la ALT	70
Conclusiones	76
Trabajo futuro	78
Apéndice A	79

Índice de tablas

2.1. Tasas de uso simuladas del componente A en los hogares y sus probabilidades correspondientes.	34
2.2. Ciclos hasta la falla en la prueba de laboratorio (ALT) (Datos simulados). . .	35
2.3. Tiempo de vida del producto en campo (Datos simulados).	36
2.4. Parámetros estimados vía Evolución Diferencial para los datos de vida del componente A sin censura.	38
2.5. Ciclos hasta la falla en la ALT que determinan la censura	39
2.6. Tiempos de censura en campo	41
2.7. Parámetros estimados vía Evolución Diferencial para los datos de vida del componente A censurados.	42
2.8. Intervalos de predicción bilaterales naive sin censura	44
2.9. Límites de predicción unilaterales inferiores naive sin censura	44
2.10. Límites de predicción unilaterales superiores naive sin censura	44
2.11. Intervalos de predicción bilaterales naive con Censura 1	44
2.12. Límites de predicción unilaterales inferiores naive con Censura 1	45
2.13. Límites de predicción unilaterales superiores naive con Censura 1	45
2.14. Intervalos de predicción bilaterales naive con Censura 2	45
2.15. Límites de predicción unilaterales inferiores naive con Censura 2	45
2.16. Límites de predicción unilaterales superiores naive con Censura 2	45
2.17. Intervalos de predicción bilaterales naive con Censura 3	46
2.18. Límites de predicción unilaterales inferiores naive con Censura 3	46
2.19. Límites de predicción unilaterales superiores naive con Censura 3	46
2.20. Límites de predicción unilaterales inferiores calibrados sin censura	52
2.21. Límites de predicción unilaterales superiores calibrados sin censura	52
2.22. Intervalos de predicción bilaterales calibrados sin censura	53

2.23. Límites de predicción unilaterales inferiores calibrados con Censura 1 (Percentil 95)	57
2.24. Límites de predicción unilaterales superiores calibrados con Censura 1 (Percentil 95)	57
2.25. Intervalos de predicción bilaterales calibrados con Censura 1	58
2.26. Límites de predicción unilaterales inferiores calibrados con Censura 2 (Percentil 80)	62
2.27. Límites de predicción unilaterales superiores con Censura 2 (Percentil 80)	62
2.28. Intervalos de predicción bilaterales calibrados con Censura 2	63
2.29. Límites de predicción unilaterales inferiores calibrados con Censura 3 (Percentil 70)	67
2.30. Límites de predicción unilaterales superiores calibrados con Censura 3 (Percentil 70)	67
2.31. Intervalos de predicción bilaterales calibrados con Censura 3	68
2.32. Longitud de los Intervalos de predicción bilaterales calibrados	68
3.1. Parámetros estimados vía Evolución Diferencial para los datos de vida del componente A sin censura sólo con la ALT.	70
3.2. Intervalos de predicción bilaterales naive sin censura sólo con la ALT	71
3.3. Límites de predicción unilaterales inferiores naive sin censura sólo con la ALT	71
3.4. Límites de predicción unilaterales superiores naive sin censura sólo con la ALT	72
3.5. Límites de predicción unilaterales inferiores calibrados sin censura en la ALT	74
3.6. Límites de predicción unilaterales superiores calibrados sin censura en la ALT	74
3.7. Intervalos de predicción bilaterales calibrados sin censura	74

Índice de figuras

1.1. Distribución de la tasa de uso del componente A en los hogares	9
1.2. Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 1)	26
1.3. Continuación Figura 1.2: Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 1)	27
1.4. Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 2)	28
1.5. Algoritmo para la calibración de los límites de predicción inferior y superior promediando las probabilidades de cobertura condicional (Método 3)	31
2.1. Curva de calibración del límite inferior por el Método 1 sin censura (CP=0.90)	48
2.2. Curva de calibración del límite superior por el Método 1 sin censura (CP=0.90)	48
2.3. Curva de calibración del límite inferior por el Método 1 sin censura (CP=0.95)	49
2.4. Curva de calibración del límite superior por el Método 1 sin censura (CP=0.95)	49
2.5. Curva de calibración del límite inferior por el Método 2 sin censura (CP=0.90)	49
2.6. Curva de calibración del límite superior por el Método 2 sin censura (CP=0.90)	49
2.7. Curva de calibración del límite inferior por el Método 2 sin censura (CP=0.95)	50
2.8. Curva de calibración del límite superior por el Método 2 sin censura (CP=0.95)	50
2.9. Curva de calibración del límite inferior por el Método 3 sin censura (CP=0.90)	51
2.10. Curva de calibración del límite superior por el Método 3 sin censura (CP=0.90)	51
2.11. Curva de calibración del límite inferior por el Método 3 sin censura (CP=0.95)	51
2.12. Curva de calibración del límite superior por el Método 3 sin censura (CP=0.95)	51
2.13. Curva de calibración del límite inferior por el Método 1 con Censura 1 (CP=0.90)	53
2.14. Curva de calibración del límite superior por el Método 1 con Censura 1 (CP=0.90)	53
2.15. Curva de calibración del límite inferior por el Método 1 con Censura 1 (CP=0.95)	54

2.16. Curva de calibración del límite superior por el Método 1 con Censura 1 (CP=0.95)	54
2.17. Curva de calibración del límite inferior por el Método 2 con Censura 1 (CP=0.90)	54
2.18. Curva de calibración del límite superior por el Método 2 con Censura 1 (CP=0.90)	54
2.19. Curva de calibración del límite inferior por el Método 2 con Censura 1 (CP=0.95)	55
2.20. Curva de calibración del límite superior por el Método 2 con Censura 1 (CP=0.95)	55
2.21. Curva de calibración del límite inferior por el Método 3 con Censura 1 (CP=0.90)	56
2.22. Curva de calibración del límite superior por el Método 3 con Censura 1 (CP=0.90)	56
2.23. Curva de calibración del límite inferior por el Método 3 con Censura 1 (CP=0.95)	56
2.24. Curva de calibración del límite superior por el Método 3 con Censura 1 (CP=0.95)	56
2.25. Curva de calibración del límite inferior por el Método 1 con Censura 2 (CP=0.90)	58
2.26. Curva de calibración del límite superior por el Método 1 con Censura 2 (CP=0.90)	58
2.27. Curva de calibración del límite inferior por el Método 1 con Censura 2 (CP=0.95)	59
2.28. Curva de calibración del límite superior por el Método 1 con Censura 2 (CP=0.95)	59
2.29. Curva de calibración del límite inferior por el Método 2 con Censura 2 (CP=0.90)	59
2.30. Curva de calibración del límite superior por el Método 2 con Censura 2 (CP=0.90)	59
2.31. Curva de calibración del límite inferior por el Método 2 con Censura 2 (CP=0.95)	60
2.32. Curva de calibración del límite superior por el Método 2 con Censura 2 (CP=0.95)	60
2.33. Curva de calibración del límite inferior por el Método 3 con Censura 2 (CP=0.90)	61
2.34. Curva de calibración del límite superior por el Método 3 con Censura 2 (CP=0.90)	61
2.35. Curva de calibración del límite inferior por el Método 3 con Censura 2 (CP=0.95)	61
2.36. Curva de calibración del límite superior por el Método 3 con Censura 2 (CP=0.95)	61
2.37. Curva de calibración del límite inferior por el Método 1 con Censura 3 (CP=0.90)	63
2.38. Curva de calibración del límite superior por el Método 1 con Censura 3 (CP=0.90)	63
2.39. Curva de calibración del límite inferior por el Método 1 con Censura 3 (CP=0.95)	64
2.40. Curva de calibración del límite superior por el Método 1 con Censura 3 (CP=0.95)	64
2.41. Curva de calibración del límite inferior por el Método 2 con Censura 3 (CP=0.90)	64
2.42. Curva de calibración del límite superior por el Método 2 con Censura 3 (CP=0.90)	64
2.43. Curva de calibración del límite inferior por el Método 2 con Censura 3 (CP=0.95)	65
2.44. Curva de calibración del límite superior por el Método 2 con Censura 3 (CP=0.95)	65
2.45. Curva de calibración del límite inferior por el Método 3 con Censura 3 (CP=0.90)	66
2.46. Curva de calibración del límite superior por el Método 3 con Censura 3 (CP=0.90)	66
2.47. Curva de calibración del límite inferior por el Método 3 con Censura 3 (CP=0.95)	66
2.48. Curva de calibración del límite superior por el Método 3 con Censura 3 (CP=0.95)	66

-
- 3.1. Curva de calibración del límite inferior por el Método 1 sin censura en la ALT (CP=0.90) 72
 - 3.2. Curva de calibración del límite superior por el Método 1 sin censura en la ALT (CP=0.90) 72
 - 3.3. Curva de calibración del límite inferior por el Método 1 sin censura en la ALT (CP=0.95) 73
 - 3.4. Curva de calibración del límite superior por el Método 1 sin censura en la ALT (CP=0.95) 73

Introducción

En la determinación del desempeño de un producto en manos del consumidor a partir de los resultados de pruebas de laboratorio, no siempre se logran identificar todas las variables asociadas, así como los diferentes modos de falla. Esta deficiencia suele reflejarse en un número de devoluciones por garantía mayor al esperado, lo cual indica que las variables evaluadas en la etapa de diseño no fueron suficientes para estimar adecuadamente dicha distribución de tiempo de falla. Es por esto que Meeker *et al.* (2009), proponen un modelo de tasa de uso para confiabilidad en campo, el cual integra los datos obtenidos a partir de las pruebas de vida acelerada (ALT) con la información complementaria de las reclamaciones por garantía y de las condiciones de uso del producto, para describir la distribución del tiempo de vida del producto en su campo de acción.

El problema a abordar en este trabajo de investigación, es la predicción estadística del tiempo de vida de productos en operación a través de la implementación del modelo de tasa de uso planteado por Meeker *et al.* (2009). Para el cumplimiento de este objetivo, se utilizará la metodología naive para generar intervalos de predicción para datos de vida, así como los procedimientos de calibración para límites de predicción naive propuestos por Meeker y Escobar (1998) y Escobar y Meeker (1999), por lo tanto será posible realizar comparaciones acerca de los métodos por medio de los cuales se obtienen los mejores resultados.

Claramente el contexto de esta investigación se desarrolla en el área de la confiabilidad, es por esto que resulta importante realizar un breve recorrido por los estudios relacionados con nuestro tema de interés. La confiabilidad aplicada emergió al final de la década de los años cuarenta (1941-1950) después de la segunda guerra mundial. Parte de sus metodologías fueron desarrolladas para estimar el número de repuestos necesarios para mantener equipos electrónicos y mecánicos funcionando en forma intensiva por periodos largos de tiempo durante la guerra. A partir de entonces, esta rama de la ingeniería se ha visto enriquecida por los valiosos aportes que las herramientas estadísticas le han proporcionado para la evaluación, la mejora, el diseño y el mantenimiento de la confiabilidad. Sin embargo, aunque el área de la confiabilidad y sus aplicaciones, como lo son las pruebas de vida acelerada, han despertado el interés científico, aún es un campo en el cual falta mucho por explorar y por tanto existe gran oportunidad para la investigación (Escobar *et al.* (2003)).

Para citar algunos de los trabajos realizados en cuanto a las pruebas de vida acelerada específicamente, como ya se ha mencionado, Meeker *et al.* (2009) proponen un modelo donde se relacionan los resultados obtenidos a partir de un ALT y la información de tasa de uso del producto en campo, además desarrollan un modelo con estas mismas características

pero aplicado a dos modos de falla. Nelson (2001) describe una aplicación y un modelo para el uso de los datos de un ALT para predecir la confiabilidad en campo de una población de unidades que son sometidas a diferentes esfuerzos dinámicos. Yang y Zaghati (2006) desarrollan modelos de tasa de uso para pruebas de vida acelerada que pueden ser utilizados cuando la distribución de los ciclos para la falla depende de la tasa de uso del producto en su campo de acción. Mannaa *et al.* (2007) ilustran las aplicaciones de un modelo de tasa de uso para predecir las utilidades obtenidas a partir de una garantía de dos dimensiones.

Una vez revisados los temas pertinentes a lo que a confiabilidad y pruebas de vida acelerada se refiere, es necesario reseñar los estudios realizados en el área de la predicción estadística para el tipo de datos obtenidos a partir de las pruebas de confiabilidad ya descritas, como son los intervalos de predicción basados en datos de vida censurados, los cuales son abordados por Meeker y Escobar (1998) y más ampliamente desarrollados en Escobar y Meeker (1999), en este último, el método general basado de una distribución paramétrica supuesta utiliza calibración por simulación. Antle y Rademaker (1972) y Nelson y Schmee (1981), desarrollan una simulación a partir de métodos de intervalos de predicción para distribuciones de localización-escala o log-localización-escala y datos con censura Tipo II. Nelson (1995) muestra un procedimiento simple para calcular los límites de predicción del número de fallas que pueden ser observadas en una inspección futura, basado en el número de fallas encontradas en una inspección previa cuando las unidades tienen una distribución Weibull de tiempo de falla. Por otro lado, Cox (1975) así como Atwood (1984), presentan aproximaciones analíticas para intervalos de predicción basados en la distribución asintótica de los estimadores de máxima verosimilitud (MLE) para los parámetros.

Este trabajo tiene la siguiente estructura. En el Capítulo 1 se desarrollan los conceptos teóricos que fundamentan la investigación como tal, es por eso que se describen las características de una prueba de vida acelerada así como los modelos de intensidad de uso de los productos en manos del cliente y los modelos de mezclas finitas, además se presenta detalladamente el modelo de tasa de uso para confiabilidad en campo propuesto por Meeker *et al.* (2009). Por otro lado, en este capítulo también se encuentran los fundamentos de la predicción estadística para datos de vida. En el Capítulo 2 se plantea la aplicación del modelo de tasa de uso a un producto que llamaremos el componente A, con base en el cual se realizará la predicción estadística, es decir, la aproximación de los intervalos y límites de predicción naive así como sus calibraciones respectivas. Para dicha aplicación, tanto la generación de los datos como la calibración se realizó vía simulación en el paquete estadístico R, por lo tanto el código implementado se presenta en el apéndice A. En el Capítulo 3 se realiza la predicción estadística para el componente A para el caso particular en el que se cuenta sólo con la información proporcionada por la prueba de laboratorio, lo cual permite hacer comparaciones sobre la pertinencia del modelo de tasa de uso para confiabilidad en campo.

Marco Teórico

1.1. Pruebas de Vida Acelerada (ALT)

Las Pruebas de Vida Acelerada (en inglés ALT) consisten en una variedad de métodos de prueba en laboratorio para acortar la vida de un producto o acelerar la degradación de su rendimiento, por lo tanto proporcionan evaluaciones oportunas de la confiabilidad de los materiales, componentes y subsistemas de los productos. El objetivo de estas pruebas es obtener datos rápidamente, los cuales modelados y analizados apropiadamente, proporcionan información deseada sobre la vida del producto o de su rendimiento bajo condiciones normales de uso (Nelson, 1990).

En las ALT se registra el tiempo en términos de una variable que es proporcional a la cantidad de uso del producto. El tiempo de uso generalmente se puede definir por los mecanismos de falla físicos fundamentales (por ejemplo, número de millas recorridas por un automóvil, número de páginas impresas por una impresora). Cualquiera que sea la definición de uso para un producto en particular, ésta debe depender del mecanismo de falla. La definición de escala de tiempo (por ejemplo, horas vs. días) es un tanto arbitraria y generalmente busca facilitar la interpretación de los resultados de una ALT (Meeker *et al.*, 2009).

1.1.1. Relación entre los resultados de la ALT y la confiabilidad en campo

Una vez realizada la prueba de laboratorio, lo interesante es poder determinar qué dicen estos resultados con respecto al funcionamiento del producto en su campo de acción, ya que debe tenerse en cuenta que las ALT son cuidadosamente controlados, mientras que el ámbito de uso normal es muy variable, por ejemplo, los productos en campo pueden tener diferentes tasas de uso promedio dentro de la población. Dadas estas condiciones, si se logra caracterizar adecuadamente el modo de uso del producto, puede ser posible utilizar los resultados de las ALT para predecir su distribución de tiempo de falla en campo. (Meeker *et al.*, 2009).

1.2. Intensidades de uso de los productos en manos del cliente y su influencia sobre las políticas de garantía

Las definiciones presentadas en esta sección son tomadas de Kim *et al.* (2001). Una garantía es un acuerdo contractual que el fabricante ofrece cuando vende su producto. Dicho acuerdo implica que el fabricante repare todas las fallas que ocurran dentro del período de garantía o que compense al comprador de alguna manera. En este contexto, la garantía sirve como un mecanismo para proteger al comprador de los productos defectuosos. Al ofrecer una garantía los fabricantes incurren en costos adicionales. Sin embargo, las garantías también sirven como señales para informar a los clientes sobre la calidad del producto. En este sentido, unas mejores condiciones en los términos de la garantía, implican una mayor calidad del producto, es por esta razón que los fabricantes tienden a utilizarlas como una herramienta de promoción para aumentar sus ventas e ingresos. Por lo tanto, desde la perspectiva del fabricante, ofrecer una garantía sólo vale la pena si los beneficios (en términos de mayores ventas o ingresos) superarán los costos adicionales asociados con el servicio de garantía.

Muchas políticas de garantía han sido propuestas y analizadas tanto para productos nuevos como usados. Estas políticas pueden agruparse en políticas unidimensionales y multidimensionales. Las políticas unidimensionales están caracterizadas por un intervalo de tiempo llamado período de garantía. El análisis de los costos de las políticas unidimensionales se basa en la suposición de que los compradores son homogéneos con respecto a la intensidad de uso (o tasa de uso) del producto. Esto implica que la tasa de uso del producto es la misma para todos los compradores. En contraste, las políticas de garantía bidimensionales están caracterizadas por una región de dos dimensiones, con un eje representando el tiempo (o la edad) y otro el uso. Estas políticas consideran que la intensidad de uso del producto en la población de compradores puede variar y se modela como una variable aleatoria en el análisis de costos.

Este trabajo centra su atención sobre las intensidades de uso del producto en manos del cliente, más que sobre las políticas de garantía, especialmente sobre las intensidades de uso que son variables dentro de la población de compradores. Un caso que ejemplifica el contexto de la tasa de uso variable, es considerar la intensidad de uso de una lavadora doméstica, en términos de carga y de frecuencia de usos por semana, la cual varía dependiendo del tamaño de la familia. Esto también es cierto para muchos otros productos domésticos e industriales. La degradación y las fallas del producto dependen de su intensidad de uso y esta a su vez tiene un impacto en el costo de la garantía prevista.

En cuanto a la modelación de la intensidad de uso, esta puede modelarse como una variable aleatoria continua o como una variable aleatoria discreta. A continuación se presentan ambos enfoques.

1.2.1. Modelación de la intensidad de uso

Sea U la variable que denota la tasa de uso. Esta es una variable aleatoria que caracteriza las diferentes tasas de uso dentro de la población. Como ya se mencionó, la tasa de uso puede modelarse como una variable aleatoria tanto continua como discreta.

Enfoque 1: Variable continua

La tasa de uso puede ser modelada como una variable aleatoria continua distribuida sobre un intervalo u_{min} y u_{max} ($u_{max} > u_{min}$) de acuerdo a una función de distribución $G(u)$ con función de densidad $g(u)$. Estos dos límites denotan las tasas de uso mínima y máxima respectivamente. Por ejemplo, dos formas de distribución que se pueden considerar son

- Distribución Uniforme: $g(u) = (1/(u_{max} - u_{min}))$
- Distribución Gamma: $g(u) = (1/\Gamma(\alpha))u^{\alpha-1}e^{-u}$ con parámetro $\alpha \geq 0$

Se puede notar que para la distribución uniforme, $u_{min} \geq 0$ y $u_{max} < \infty$ y para la distribución gamma, $u_{min} = 0$ y $u_{max} = \infty$.

Condicionado a la tasa de uso $U = u$, θ (parámetro de escala en la función de distribución de falla del producto $F(t; \theta)$) está dado por

$$\theta(u) = \delta(u)\theta_d,$$

donde θ_d es un parámetro de diseño y $\delta(u)$ define el efecto de la tasa de uso (o carga sobre la unidad) y es modelado como

$$\delta(u) = \begin{cases} 1 & \text{si } u_{min} \leq u \leq u_0 \\ (u/u_0)^\gamma & \text{si } u_0 < u \leq u_{max} \end{cases}$$

con $\gamma (\geq 1)$ y u_0 como los dos parámetros de diseño adicionales. Los parámetros de diseño dependen de las decisiones de diseño y están bajo el control del fabricante. Valores altos de θ_d y u_0 y valores pequeños de γ son el resultado de un mejor diseño.

Enfoque 2: Variable discreta

Este enfoque considera que la tasa de uso en la población puede ser agrupado dentro de varios grupos. El caso más simple es agruparlos en dos grupos, es decir, los usuarios “livianos” con tasa de uso u_l y los usuarios “pesados ” con tasa de uso u_p . Entonces, condicionado a la tasa de uso $U = u$, θ (parámetro de escala en la función de distribución de falla del producto $F(t; \theta)$) está dado por

$$\theta(u) = \phi\theta_d.$$

Para el usuario “liviano”, $\phi = 1$, y $\theta(u) = \theta_l = \theta_d$. Para el usuario “pesado”, $\phi > 1$, y $\theta(u) = \theta_p = \phi\theta_d$. Además, $p[(1 - p)]$ denota la probabilidad de que el comprador sea un usuario liviano [pesado].

Dada la posibilidad de variación en la tasa de uso promedio del producto en la población, una posible alternativa en la modelación de la distribución de tiempo de vida del producto, es considerar una distribución donde la población se divida en subpoblaciones de acuerdo a las diferentes tasas de uso, por lo tanto, a su vez, estas tasas de uso tendrían una distribución discreta, lo que implicaría considerar un modelo de mezclas finitas para el tiempo de vida. En la siguiente sección se presentan las principales características de este tipo de modelos.

1.3. Modelos de mezclas finitas

En muchas ocasiones se tiene la necesidad de estudiar modelos flexibles que puedan dar un mejor ajuste al comportamiento de diversos fenómenos aleatorios en donde existe la presencia de grupos ocultos que conducen a la heterocedasticidad y ocasionan que en la población se tengan varios máximos o distribuciones con ciertas semejanzas a las distribuciones comúnmente usadas, pero con diferencias que hacen que una sola distribución no ajuste adecuadamente los datos. Lo anterior ocasiona que un modelo propuesto para ajustar el comportamiento del fenómeno en estudio se deba proponer por medio de una mezcla de distribuciones. Actualmente estos modelos van ganando terreno en las aplicaciones que van desde las ciencias biológicas y la confiabilidad hasta las ciencias económicas (Titterton *et al.* (1985)). Las definiciones presentadas en esta sección son tomadas de McLachlan y Peel (2000).

1.3.1. Definición básica

Sea $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ una muestra aleatoria de tamaño n , donde \mathbf{Y}_j es un vector aleatorio p -dimensional con fdp $f(\mathbf{y}_j)$ de \mathbb{R}^p . En la práctica, \mathbf{Y}_j contiene las variables aleatorias correspondientes a las p mediciones hechas a los j -ésimos registros de algunas características del fenómeno bajo estudio. Sea $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_n^T)^T$, donde el superíndice T denota el vector transpuesto. Note que se usa \mathbf{Y} para representar la muestra completa; esto es, \mathbf{Y} es una n -tupla de puntos en \mathbb{R}^p . Cuando sea posible, una realización de un vector aleatorio se denota con la correspondiente letra minúscula. Por ejemplo, $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$ denota una muestra aleatoria observada cuando \mathbf{y}_j es el valor observado del vector aleatorio \mathbf{Y}_j .

Aunque aquí se tome el vector \mathbf{Y}_j como un vector aleatorio continuo, se puede seguir viendo a $f(\mathbf{y}_j)$ como una función de probabilidad de masa en el caso donde \mathbf{Y}_j sea discreto. Suponga que la densidad $f(\mathbf{y}_j)$ de \mathbf{Y}_j puede ser escrita en la forma

$$f(\mathbf{y}_j) = \sum_{i=1}^k \pi_i f_i(\mathbf{y}_j), \quad (1.1)$$

donde $f_i(\mathbf{y}_j)$ son densidades y los π_i son cantidades no negativas que suman uno; esto es

$$0 \leq \pi_i \leq 1 \quad (i = 1, \dots, k) \quad \text{y} \quad \sum_{i=1}^k \pi_i = 1.$$

Las cantidades π_1, \dots, π_k son llamadas proporciones de la mezcla o pesos. Como las funciones, $f_1(\mathbf{y}_j), \dots, f_k(\mathbf{y}_j)$, son densidades, es obvio que (1.1) define una densidad. Las $f_i(\mathbf{y}_j)$ son llamadas las densidades de los componentes de la mezcla. Entonces, (1.1) es la densidad de una mezcla finita de k -componentes y su correspondiente función de distribución $F(\mathbf{y}_j)$ es la distribución de una mezcla finita de k -componentes.

1.3.2. Interpretación de los modelos de mezcla

Una manera obvia de generar un vector aleatorio \mathbf{Y}_j con la densidad de la mezcla de k -componentes $f_i(\mathbf{y}_j)$, dado en (1.1), es como sigue. Sea Z_j una variable aleatoria categórica que toma los valores $1, \dots, k$ con probabilidades π_1, \dots, π_k , respectivamente, y suponga que la densidad condicional de \mathbf{Y}_j dado $Z_j = i$ es $f_i(\mathbf{y}_j)$ ($i = 1, \dots, k$). Luego la densidad incondicional de \mathbf{Y}_j (es decir, su densidad marginal) está dada por $f(\mathbf{y}_j)$. En este contexto, la variable Z_j se puede considerar como la etiqueta de la componente del vector \mathbf{Y}_j . Es conveniente trabajar con el vector k -dimensional \mathbf{Z}_j en lugar de la variable categórica única Z_j , donde el i -ésimo elemento, Z_{ij} , de \mathbf{Z}_j , toma los valores de uno o cero, en función de si la componente de origen de \mathbf{Y}_j en la mezcla es igual a i o no ($i = 1, \dots, k$). Por lo tanto \mathbf{Z}_j se distribuye según una distribución multinomial de k categorías con probabilidades π_1, \dots, π_k ; esto es

$$\Pr\{\mathbf{Z}_j = \mathbf{z}_j\} = \pi_1^{z_{1j}} \pi_2^{z_{2j}} \dots \pi_k^{z_{kj}}.$$

Entonces,

$$\mathbf{Z}_j \sim \text{Multinomial}_k(\mathbf{1}, \boldsymbol{\pi}),$$

donde $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)^T$.

De acuerdo a la interpretación dada sobre un modelo de mezcla, una situación obvia donde el modelo de mezcla de k -componentes (1.1) es directamente aplicable, es cuando \mathbf{Y}_j se extrae de una población K que consta de k grupos, K_1, \dots, K_k , en proporciones π_1, \dots, π_k . Si la densidad de \mathbf{Y}_j en el grupo K_i está dada por $f_i(\mathbf{y}_j)$ para ($i = 1, \dots, k$), entonces la densidad de \mathbf{Y}_j tiene la forma de mezcla de k -componentes (1.1). En esta situación, las k componentes de la mezcla pueden ser identificadas físicamente con los k grupos externos existentes, K_1, \dots, K_k .

1.3.3. Estructura de datos incompletos en el problema de mezcla

Previamente se introdujo la definición de el vector de etiquetas \mathbf{Z}_j con variables indicadoras cero-uno que definen la componente en el modelo de mezcla (1.1) a partir de la cual se considera que el vector aleatorio \mathbf{Y}_j ha surgido. Esta conceptualización del modelo de mezcla en términos de \mathbf{Y}_j y \mathbf{Z}_j es muy útil ya que permite la estimación por máxima verosimilitud de la distribución de la mezcla que puede ser calculada a través de diferentes metodologías.

La estimación de la distribución de la mezcla se hace con base en los datos $\mathbf{y}_1, \dots, \mathbf{y}_n$, usualmente disponibles en forma de una muestra aleatoria observada tomada de la densidad de la mezcla (1.1). Esto es, $\mathbf{y}_1, \dots, \mathbf{y}_n$ son los valores realizados de los n independientes e idénticamente distribuidos (i.i.d.) vectores aleatorios $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ con densidad común $f(\mathbf{y}_j)$. Escribimos

$$\mathbf{Y}_1, \dots, \mathbf{Y}_n \stackrel{i.i.d.}{\sim} F,$$

donde $F(\mathbf{y}_j)$ denota la distribución correspondiente a la densidad de la mezcla $f(\mathbf{y}_j)$.

En este enfoque, los datos $\mathbf{y}_1, \dots, \mathbf{y}_n$ son vistos como incompletos ya que sus vectores indicadores asociados, $\mathbf{z}_1, \dots, \mathbf{z}_n$, no están disponibles. Por lo tanto, el vector de datos completos es

$$\mathbf{y}_c = (\mathbf{y}^T, \mathbf{z}^T)^T, \quad (1.2)$$

donde

$$\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T,$$

es el vector de datos observados o de datos incompletos y donde

$$\mathbf{z} = (\mathbf{z}_1^T, \dots, \mathbf{z}_n^T)^T,$$

es el vector no observable de variables indicadoras.

Los vectores de etiquetas $\mathbf{z}_1, \dots, \mathbf{z}_n$ se consideran como los valores realizados de los vectores aleatorios $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, donde, para datos independientes, es apropiado asumir que se distribuyen incondicionalmente como

$$\mathbf{Z}_1, \dots, \mathbf{Z}_n \stackrel{i.i.d.}{\sim} \text{Multinomial}_k(1, \boldsymbol{\pi}). \quad (1.3)$$

Se puede observar que en este contexto de datos incompletos, el modelo de mezcla surge porque los vectores de etiquetas son vectores faltantes del vector de datos completos y tenemos que estimar la distribución de la mezcla sobre los datos disponibles provenientes de la distribución marginal de \mathbf{Y}_j en lugar de a partir de la distribución conjunta del vector \mathbf{Y}_j y su vector de etiquetas \mathbf{Z}_j .

Si el vector de datos completos \mathbf{y}_c estuviera disponible, la estimación de la distribución de la mezcla sería más sencilla que sobre la base de los datos observados \mathbf{y} , puesto que cada componente de densidad $f_i(\mathbf{y})$ podría ser estimada directamente a partir de los datos que se sabe han venido de ella; es decir, a partir de los datos \mathbf{y}_j con $z_{ij} = 1$. Luego, los otros parámetros a ser estimados serían las proporciones de la mezcla, las cuales, en el caso de un diseño muestral para los datos clasificados, se pueden estimar por la proporción de los datos de cada componente, a saber

$$\hat{\pi}_i = \sum_{j=1}^n z_{ij}/n \quad (i = 1, \dots, k).$$

1.3.4. Formulación paramétrica del modelo de mezcla

En muchas aplicaciones, las densidades de las componentes $f_i(\mathbf{y}_j)$ son especificadas para pertenecer a alguna familia paramétrica. En este caso, las densidades de las componentes $f_i(\mathbf{y}_j)$ se especifican como $f_i(\mathbf{y}_j; \boldsymbol{\theta}_i)$, donde $\boldsymbol{\theta}_i$ es el vector de parámetros desconocidos

en la forma postulada para la i -ésima densidad en la mezcla. Entonces, la densidad de la mezcla $f(\mathbf{y}_j)$ puede escribirse como

$$f(\mathbf{y}_j; \Psi) = \sum_{i=1}^k \pi_i f_i(\mathbf{y}_j; \theta_i), \quad (1.4)$$

donde el vector Ψ que contiene todos los parámetros desconocidos en el modelo de mezcla puede escribirse como

$$\Psi = (\pi_1, \dots, \pi_{k-1}, \xi^T)^T, \quad (1.5)$$

donde ξ es el vector que contiene todos los parámetros en $\theta_1, \dots, \theta_k$ conocido *a priori*. Dejamos que Ω denote el espacio paramétrico especificado para Ψ . Desde que las proporciones de la mezcla π_i suman uno, una de estas proporciones es redundante. En la definición de Ψ en (1.5), se ha omitido arbitrariamente la k -ésima proporción de la mezcla π_k .

1.3.5. Estimación por máxima verosimilitud para el modelo de mezcla

Con la aproximación por Máxima Verosimilitud (ML, por sus siglas en inglés) para la estimación de un vector de parámetro d -dimensional Ψ en una densidad postulada $f(\mathbf{y}_j; \Psi)$ para el vector aleatorio \mathbf{Y}_j asociado con el j -ésimo registro del fenómeno bajo estudio, un estimador $\hat{\Psi}$ es proporcionado en situaciones regulares, por una solución apropiada de la ecuación de verosimilitud,

$$\partial L(\Psi)/\partial(\Psi) = \mathbf{0},$$

o equivalentemente,

$$\partial \log L(\Psi)/\partial(\Psi) = \mathbf{0},$$

donde

$$\begin{aligned} L(\Psi) &= \prod_{j=1}^n f(\mathbf{y}_j; \Psi) \\ &= \prod_{j=1}^n \left(\sum_{i=1}^k \pi_i f_i(\mathbf{y}_j; \theta_i) \right), \end{aligned} \quad (1.6)$$

denota la función de verosimilitud para Ψ bajo el supuesto de independencia de la muestra aleatoria observada, $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$. O equivalentemente,

$$\begin{aligned}\log L(\Psi) &= \sum_{j=1}^n \log f(\mathbf{y}_j; \Psi) \\ &= \sum_{j=1}^n \log \left\{ \sum_{i=1}^k \pi_i f_i(\mathbf{y}_j; \theta_i) \right\},\end{aligned}\tag{1.7}$$

denota la función de log verosimilitud para Ψ bajo el supuesto de independencia de la muestra aleatoria observada, $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$.

Como se mencionó anteriormente, dentro de la formulación del problema de mezcla en el enfoque de datos incompletos, el vector de datos observados

$$\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$$

es visto como incompleto cuando los vectores de etiquetas, $\mathbf{z}_1, \dots, \mathbf{z}_n$, no están disponibles. Según esta estructura, donde cada \mathbf{y}_j proviene de una de las componentes del modelo de mezcla (1.4) que ha sido ajustado, \mathbf{z}_j es un vector k -dimensional con $z_{ij} = (\mathbf{z}_j)_i = 1$ o 0 , de acuerdo a si \mathbf{y}_j proviene o no de la i -ésima componente de la mezcla ($i = 1, \dots, k; j = 1, \dots, n$). Teniendo en cuenta estas características, el log de la verosimilitud de los datos completos en (1.3) para Ψ , está dado por

$$\log L_c(\Psi) = \sum_{i=1}^k \sum_{j=1}^n z_{ij} \{ \log \pi_i + \log f_i(\mathbf{y}_j; \theta_i) \}.\tag{1.8}$$

1.4. El modelo de tasa de uso para confiabilidad en campo

Para los casos en los cuales no está disponible la información que permita caracterizar en laboratorio las condiciones de uso del producto en campo, pero tanto los datos de pruebas de vida y los datos de campo (procedentes, por ejemplo, de las devoluciones por garantía) están disponibles, Meeker *et al.* (2009) proponen un modelo para relacionar los dos conjuntos de datos. Bajo un conjunto razonable de supuestos prácticos, este modelo puede ser utilizado para predecir la distribución del tiempo de falla para un componente futuro o para un producto operando bajo las mismas condiciones de uso.

Dado que el objetivo de este trabajo de investigación es la construcción de intervalos de predicción para el tiempo de vida de un producto con base en el modelo de tasa de uso para confiabilidad en campo propuesto por Meeker *et al.* (2009), se partirá de un problema con el componente A que ellos mismos plantean en su artículo, el cual se replicará vía simulación con el fin de obtener los datos que permitan construir dichos intervalos. El problema con el componente A se describe a continuación.

Un electrodoméstico estaba experimentando una tasa de devoluciones por garantía mayor a la esperada. Se descubrió que la causa de la gran mayoría de las devoluciones fue un defecto de diseño en un componente en particular que llamaremos componente A. Se realizó un

cambio en el diseño que espera alargar la vida útil del componente A y se llevó a cabo una ALT para estimar el tiempo de vida de las unidades fabricadas bajo el nuevo diseño. Los ingenieros responsables de la confiabilidad de este producto proporcionaron la siguiente información:

- El modo de falla visto en campo puede ser reproducido con exactitud en una ALT realizado con una tasa de ciclos aumentada y siguiendo muy de cerca las condiciones de funcionamiento del componente cuando se instala en el aparato.
- La tasa de ciclos aumentada fue elegida lo suficientemente alta para obtener resultados de la prueba en un tiempo razonable, pero no tan alta como para cambiar la distribución de los ciclos hasta la falla del componente A. En particular, a el componente A se le dio suficiente tiempo de “enfriamiento” entre cada ciclo de prueba.
- La distribución de los ciclos hasta la falla puede ser estimada con buena precisión a partir de un número de unidades en prueba relativamente pequeño, porque bajo condiciones de laboratorio cuidadosamente controladas, hubo variabilidad relativamente pequeña en el número de ciclos necesarios para hacer que el componente A fallara (cuando se compara con la variabilidad en los datos de semanas hasta la falla provenientes de campo).
- El número promedio de usos por semana del aparato varió entre los hogares de acuerdo con una distribución que fue estimada a partir de los datos de una encuesta obtenidos por el departamento de marketing de la compañía. En la Figura 1.1 se muestra el histograma de la distribución estimada de los ciclos de uso por semana en los hogares. Esta es una distribución lognormal discretizada, truncada a los 20 usos por semana; el número de hogares con más de 20 usos por semana fue despreciable.

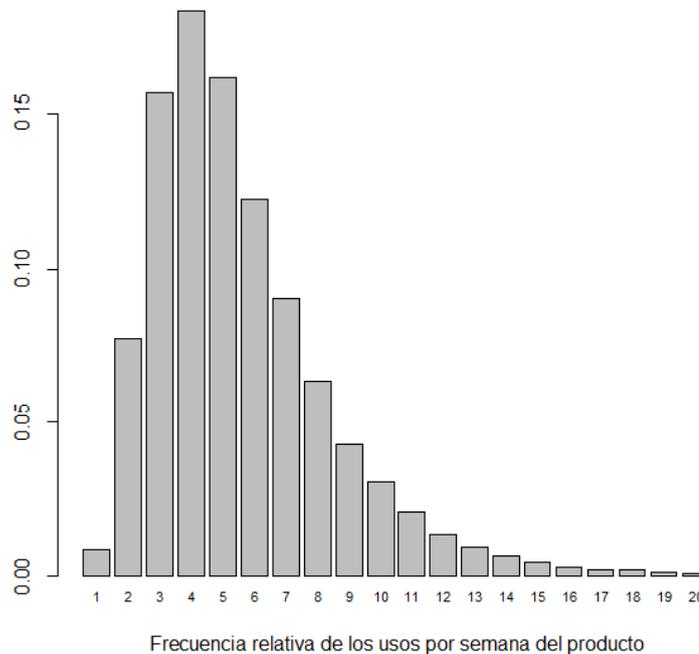


Figura 1.1: Distribución de la tasa de uso del componente A en los hogares

Meeker *et al.* (2009) parten del supuesto de que la distribución de los ciclos, C , hasta la falla del producto dada por la ALT es del tipo log-localización-escala:

$$F_C(c) = \Pr(C \leq c) = \Phi \left[\frac{\log(c/\eta_C)}{\sigma_C} \right] = \Phi \left[\frac{\log(c) - \log(\eta_C)}{\sigma_C} \right], \quad (1.9)$$

donde Φ es la cdf de una distribución completamente especificada, η_C es un parámetro de escala dado en las unidades usadas (generalmente en ciclos) en la ALT. Por ejemplo, el parámetro η_C es la mediana para la distribución lognormal y aproximadamente el cuantil 0.63 para la distribución Weibull, y σ_C es un parámetro de forma o dispersión. En la distribución Weibull, el recíproco de σ_C es usualmente el parámetro de forma.

Además, Meeker *et al.* (2009), suponen también que la distribución de la tasa de uso promedio R (ciclos por unidad de tiempo, por ejemplo, usos por semana) en manos del cliente puede ser descrita por una distribución discreta con posibles valores:

$$\mathbf{R} = (R_1, R_2, \dots, R_k),$$

con probabilidades correspondientes:

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k), \text{ donde } \sum_{i=1}^k \pi_i = 1.$$

Si los ciclos de prueba en la ALT tienen el mismo efecto sobre el tiempo de vida que los ciclos de uso del producto en servicio, entonces la fracción de falla en el tiempo real t en la población del producto (por ejemplo semanas de servicio) puede ser descrita por la siguiente distribución de mezcla finita:

$$F(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \Pr(T \leq t) = \sum_{i=1}^k \pi_i \Phi \left[\frac{\log(R_i \times t/\eta_C)}{\sigma_C} \right]. \quad (1.10)$$

Según Meeker *et al.* (2009), también es posible usar una distribución de mezcla continua para describir un modelo de tasa de uso, pero argumentan que la mezcla discreta es especialmente útil para explicar el concepto a ingenieros u otro tipo de profesionales que no estén familiarizados con distribuciones de mezcla continuas.

En la siguiente subsección se presentan los procedimientos de estimación para el modelo de tasa de uso bajo dos escenarios, para datos de vida sin censura y con censura. Cabe anotar que el caso con censura representa mejor la realidad de los productos que salen al mercado, sin embargo, se quisieron tener en cuenta ambos casos para realizar una comparación de los resultados.

Antes de presentar dichos procedimientos, es necesario definir el concepto de censura para datos de vida. Se tienen datos de vida censurados cuando no se conocen los tiempos de falla de ciertas unidades de manera exacta, solo se conocen los intervalos de tiempo donde ocurrieron o podrían haber ocurrido las fallas (Meeker y Escobar (1998)). De acuerdo a sus características específicas, se pueden tener diferentes clasificaciones y tipos de censura, estas son

- **Censura a derecha (tipo I y II):** La tipo I ocurre cuando se tienen unidades sin fallar al limitar el tiempo de observación. Cuando se limita el tiempo hasta que fallan r unidades, se tiene censura tipo II para las $(n - r)$ unidades sobrevivientes.
- **Censura a izquierda:** Ocurre cuando al inspeccionar las unidades después de un periodo de tiempo se encuentra que algunas fallaron, pero no se sabe el momento exacto en que sucedió.
- **Censura por intervalo:** Ocurre cuando se inspecciona en intervalos de tiempo y se observan fallas en cierto intervalo pero no se conoce exactamente en qué momento sucedieron, se censuran los productos sobrevivientes.
- **Censura múltiple:** Cuando en el mismo estudio se tienen diferentes tipos de censura.
- **Censura aleatoria:** Ocurre cuando cada unidad tiene un tiempo de censura que es estadísticamente independiente de su tiempo de falla, por lo tanto, el tiempo de censura no está predeterminado sino que es aleatorio. El valor observado es el mínimo entre el tiempos de censura y el tiempo de falla; las unidades cuyos tiempos de falla sean mayores que sus tiempos de censura, son censurados a derecha. En el caso de la prueba, se dice que la censura aleatoria ocurre cuando alguna unidad falla por causas ajenas a la misma.

1.4.1. Estimación por máxima verosimilitud para el modelo de tasa de uso con datos de vida sin censura

Como ya se ha mencionado, las ALT se realizan cuando los fabricantes desean tener resultados de confiabilidad para sus productos más rápidamente que bajo condiciones de funcionamiento normal, es por eso que se tratan de acelerar las fallas sometiendo los productos a niveles altos de esfuerzo para después extrapolar la confiabilidad a condiciones normales de operación (Meeker y Escobar (1998)). Tener datos de vida no censurados en la ALT, como en este caso, quiere decir que la prueba hecha a los productos en el laboratorio se realizó hasta que todas las unidades fallaron sin fijar un tiempo límite de observación.

La expresión para la función de densidad (fdp) para los ciclos hasta la falla determinada a partir de una prueba de laboratorio (ALT) está dada por

$$\begin{aligned}
 f_{\text{ALT}}(c; \eta_C, \sigma_C) &= \frac{\partial F_C(c; \eta_C, \sigma_C)}{\partial c} \\
 &= \frac{1}{c \sigma_C} \phi \left[\frac{\log(c) - \log(\eta_C)}{\sigma_C} \right], \tag{1.11}
 \end{aligned}$$

donde ϕ es una fdp. completamente especificada, η_C es un parámetro de escala dado en unidades de los ciclos en la ALT y σ_C es un parámetro de forma o dispersión (Meeker *et al.* (2009)).

La función de verosimilitud para los datos sin censura obtenidos a partir de la ALT puede expresarse como

$$L_{\text{ALT}}(c; \eta_C, \sigma_C) = \prod_{j=1}^m \frac{1}{c_j \sigma_C} \phi \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right].$$

Ahora, la función de log verosimilitud para los datos sin censura obtenidos a partir de la ALT es

$$\mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C) = \sum_{j=1}^m \left\{ -\log(c_j) - \log(\sigma_C) + \log \left(\phi \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right) \right\}.$$

Por otro lado, cuando las unidades están en campo, al fabricante le interesa determinar las unidades que han fallado durante el periodo de garantía que él ofreció a sus clientes. Si recibe reclamos por garantías de todas las unidades que salieron al mismo tiempo al mercado, quiere decir que ninguna quedará censurada ya que se conocen los tiempos exactos de falla para cada una de ellas.

Entonces, la función de densidad (fdp) para la mezcla finita que describe la fracción de falla en el tiempo real t para el producto en campo, en la escala de tiempo de servicio, está dada por

$$\begin{aligned} f_{\text{Campo}}(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) &= \frac{\partial F(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R})}{\partial t} \\ &= \sum_{i=1}^k \frac{\pi_i}{t \sigma_C} \phi \left[\frac{\log(R_i \times t) - \log(\eta_C)}{\sigma_C} \right], \end{aligned} \quad (1.12)$$

donde cada valor de R_i representa la tasa de uso promedio del producto en manos del cliente en unidades de ciclos por unidad de tiempo y cada π es su probabilidad correspondiente (Meeker *et al.* (2009)).

Los datos completos categorizados pueden ser expresados como

$$\mathbf{t}_{\text{comp}} = (\mathbf{t}^T, \mathbf{z}^T)^T$$

donde

$$\mathbf{z} = (\mathbf{z}_1^T, \dots, \mathbf{z}_n^T)^T,$$

cada $\mathbf{z}_j = (z_{ij}, i = 1, \dots, k)$ es un vector indicador de magnitud k con 1 en la posición correspondiente a la categoría apropiada y 0 en caso contrario (Titterton *et al.* (1985)).

Entonces, la función de verosimilitud en campo para los datos completos sin censura puede ser escrita en la forma

$$L_{\text{Campo}}(\mathbf{t}_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \prod_{j=1}^n \sum_{i=1}^k \pi_i^{z_{ij}} \left(\phi \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}}.$$

Luego, la función de log verosimilitud para los datos en campo sin censura es

$$\mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \sum_{j=1}^n \log \left\{ \sum_{i=1}^k \pi_i^{z_{ij}} \left(\phi \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}} \right\}.$$

Ahora, el log de la verosimilitud total para los datos de campo y los datos en la ALT sin censura está dada por

$$\mathcal{L}(t_{\text{comp}}, c; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) + \mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C). \quad (1.13)$$

Esta función de log verosimilitud total representa la probabilidad conjunta de los datos, asumiendo que las n y m observaciones son independientes (Meeker y Escobar (1998)). Usualmente no es posible realizar el cálculo directo de los MLE bajo este esquema, sin embargo, cuando los valores de los Z_{ij} y los π_i son conocidos previamente, como ocurre en nuestro caso, debido a que se tiene la información respecto a las tasas de uso que dan origen a la mezcla y por lo tanto el vector \mathbf{Z}_j es conocido para cada caso, no es necesario recurrir a otros métodos como el procedimiento iterativo conocido como algoritmo EM (Maximización del Valor Esperado), que es el que se utiliza frecuentemente para solucionar problemas de estimación bajo la interpretación de los datos de la mezcla como datos incompletos (Titterton *et al.* (1985)).

1.4.2. Estimación por máxima verosimilitud para el modelo de tasa de uso con datos de vida censurados

Durante la ejecución de una ALT, puede suceder que el tiempo para observar las fallas sea muy largo por lo que resulta necesario cortarlo o limitarlo dando lugar a observaciones censuradas. Este caso es el segundo escenario que se va a considerar para el modelo de tasa de uso.

Debido a que las unidades de la ALT están dadas en ciclos hasta la falla y teniendo en cuenta las características de censura expuestas previamente, en la prueba de laboratorio para la aplicación del modelo de tasa de uso se tiene una censura a derecha no aleatoria, esto se dá porque la prueba se realiza hasta un número de ciclos predeterminado, por lo tanto el número de unidades que no hayan fallado hasta este número de ciclos hasta la falla quedarán censuradas.

Entonces, la función de verosimilitud para los datos censurados obtenidos a partir de la ALT y que lleva información de observaciones exactas (no censuradas) y de observaciones con censura a derecha no aleatoria (Meeker y Escobar (1998) y Meeker *et al.* (2009)), puede expresarse como

$$L_{\text{ALT}}(c; \eta_C, \sigma_C) = \prod_{j=1}^m \left\{ \left(\frac{1}{c_j \sigma_C} \phi \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{\tau_j} \times \left(1 - \Phi \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{1 - \tau_j} \right\},$$

donde τ_j es el indicador de censura que toma valores de $\tau_j = 1$ si c_j es una observación exacta en la ALT y $\tau_j = 0$ si c_j es una observación censurada a derecha (ciclos hasta el fin de la prueba sin observar falla) en la ALT (Meeker y Escobar (1998) y Meeker *et al.* (2009)).

Luego, la función de log verosimilitud para los datos censurados obtenidos a partir de la ALT es

$$\begin{aligned} \mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C) &= \sum_{j=1}^m \tau_j \left\{ -\log(c_j) - \log(\sigma_C) + \log \left(\phi \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right) \right\} \\ &+ (1 - \tau_j) \left\{ \log \left(1 - \Phi \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right) \right\}. \end{aligned}$$

Por otro lado, los datos de retornos por garantía son censurados cuando no todas las unidades que están en manos del cliente han fallado durante este periodo de tiempo. Es decir que en este segundo escenario, las unidades en campo tienen una censura a derecha no aleatoria.

Nuevamente, los datos completos categorizados pueden ser expresados como

$$\mathbf{t}_{\text{comp}} = (\mathbf{t}^T, \mathbf{z}^T)^T$$

donde

$$\mathbf{z} = (\mathbf{z}_1^T, \dots, \mathbf{z}_n^T)^T,$$

Entonces, la función de verosimilitud en campo para los datos completos censurados que lleva información de observaciones exactas (no censuradas) y de observaciones con censura a derecha no aleatoria (Meeker y Escobar (1998) y Meeker *et al.* (2009)), puede ser escrita en la forma

$$\begin{aligned} L_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) &= \prod_{j=1}^n \left\{ \sum_{i=1}^k \pi_i^{z_{ij}} \left(\phi \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}} \right\}^{\delta_j} \\ &\times \left\{ 1 - \sum_{i=1}^k \pi_i^{z_{ij}} \left(\Phi \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{z_{ij}} \right\}^{1-\delta_j}, \end{aligned}$$

donde δ_j es el indicador de censura que toma valores de $\delta_j = 1$ si t_j es una observación exacta en campo y $\delta_j = 0$ si t_j es una observación censurada a derecha (tiempos de falla no observados) en campo (Meeker y Escobar (1998) y Meeker *et al.* (2009)).

Luego, la función de log verosimilitud para los datos en campo censurados es

$$\begin{aligned} \mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) &= \sum_{j=1}^n \delta_j \log \left\{ \sum_{i=1}^k \pi_i^{z_{ij}} \left(\phi \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}} \right\} \\ &+ (1 - \delta_j) \log \left\{ 1 - \sum_{i=1}^k \pi_i^{z_{ij}} \left(\Phi \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{z_{ij}} \right\}. \end{aligned}$$

Ahora, la log verosimilitud total para los datos de campo y los datos en la ALT censurados está dada por

$$\mathcal{L}(t_{\text{comp}}, c; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) + \mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C). \quad (1.14)$$

Esta función de log verosimilitud total representa la probabilidad conjunta de los datos censurados, asumiendo que las n y m observaciones son independientes (Meeker y Escobar (1998)). En este caso se puede realizar el cálculo directo de los MLE, debido a que los valores de los π_i y los Z_{ij} son conocidos previamente.

Expuestas en las secciones anteriores las características específicas del modelo de tasa de uso para el cual se pretenden construir los intervalos de predicción, resta presentar lo concerniente a la predicción estadística. Por lo tanto, en las siguientes dos secciones se desarrollan ampliamente los conceptos relacionados con este tema.

1.5. Predicción estadística

Las definiciones presentadas en esta sección son tomadas de Fredette (2004). En el análisis estadístico tradicional, se usa la información contenida en una muestra para hacer inferencias sobre la población de la cual proviene. Usualmente estas inferencias se basan en estimaciones, intervalos de confianza o en pruebas de hipótesis para los parámetros de un modelo específico. Si este modelo describe adecuadamente la población, el análisis derivado de dichas inferencias es el apropiado para la mayoría de los problemas científicos. Por otro lado, también se pueden encontrar problemas en los cuales el entendimiento del comportamiento de la población por sí mismo no resulta ser de interés, si no más bien un recurso a partir del cual se pueden predecir eventos futuros. A esta clase de problemas se le llaman problemas de predicción. Dado que los problemas de predicción son rara vez un caso de deducción lógica, el uso de herramientas probabilísticas y estadísticas son inevitables en cualquier aproximación científica que pretenda resolverlos.

Se considera un marco general paramétrico en el que la predicción se basa en una familia de modelos de la forma

$$F(y|x; \boldsymbol{\theta}) = \Pr(\mathbf{Y} \leq y | \mathbf{X} = x; \boldsymbol{\theta}),$$

donde \mathbf{X} e \mathbf{Y} son vectores de variables aleatorias que representan datos pasados y futuros respectivamente. La función de distribución F es especificada por el parámetro desconocido

p -dimensional θ , el cual puede ser estimado a partir de los valores de \mathbf{X} . Típicamente, la variable \mathbf{X} es multivariada y la variable \mathbf{Y} se asume univariada, aunque también puede ser una función de resultados multivariados, por ejemplo una suma, una media o un estadístico de orden (Lawless y Fredette (2005)).

1.5.1. Intervalos de predicción

Cuando se quiere inferir acerca de un parámetro desconocido, dos aproximaciones populares consisten en encontrar una estimación puntual o un intervalo de posibles valores para este parámetro. En un problema de predicción, se procede de manera similar cuando se pretende predecir el valor futuro de la variable aleatoria \mathbf{Y} a partir del conjunto de variables aleatorias $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de cierta población, es decir, se puede tratar de encontrar un predictor puntual o un intervalo de predicción. Siguiendo nuestro interés, se enfocará el estudio en los intervalos de predicción como tipo particular de predicción.

Definición: Un intervalo de predicción (PI por sus siglas en inglés), es un intervalo que usa la información de una muestra pasada para predecir, con una probabilidad específica, los resultados de una muestra futura de la misma población. Un intervalo con límites inferior y superior $L(\mathbf{X})$ y $U(\mathbf{X})$ respectivamente, es llamado un **intervalo de predicción bilateral** $(1 - \alpha)$ para \mathbf{Y} si

$$P[L(\mathbf{X}) \leq \mathbf{Y} \leq U(\mathbf{X})] = 1 - \alpha. \quad (1.15)$$

Cuando $\dim(\mathbf{Y}) > 1$, una región $R(\mathbf{X})$ tal que:

$$P[\mathbf{Y} \in R(\mathbf{X})] = 1 - \alpha$$

es llamada una **región de predicción** $(1 - \alpha)$ para el vector \mathbf{Y} . La cantidad $(1 - \alpha)$ es llamada la **probabilidad de cobertura** del intervalo de predicción o región de predicción.

Esta definición significa que, bajo muestreo repetido de \mathbf{X} e \mathbf{Y} , y estará incluida en $[L(\mathbf{X}), U(\mathbf{X})]$ ó $R(\mathbf{X})$, en $(1 - \alpha)100\%$ de las muestras.

En general, la distribución de las tres variables aleatorias $L(\mathbf{X})$, \mathbf{Y} y $U(\mathbf{X})$ dependerá de un parámetro desconocido θ . Por lo tanto, puede no ser posible que (1.15) sea independiente de θ . En este caso, se llamará a $[L(\mathbf{X}), U(\mathbf{X})]$ un intervalo de predicción bilateral $(1 - \alpha)$ para \mathbf{Y} si

$$P[L(\mathbf{X}) \leq \mathbf{Y} \leq U(\mathbf{X}); \theta] \geq 1 - \alpha,$$

para todo θ .

1.5.2. Intervalos de predicción bajo el marco frecuentista

Como se mencionó anteriormente, en algunas ocasiones no es posible encontrar un intervalo de predicción $(1 - \alpha)$ con una probabilidad de cobertura independiente del parámetro

desconocido θ , por lo tanto, se debe encontrar un procedimiento tal que la probabilidad de cobertura nunca sea menor que $(1 - \alpha)$ para todos los posibles valores de θ . Desafortunadamente, esta aproximación con frecuencia provee intervalos de predicción triviales (es decir, intervalos que incluyen todos los posibles valores de \mathbf{Y}).

Bajo la visión frecuentista (es decir, cuando θ es desconocido, pero se asume constante), existe una exhaustiva lista de métodos para encontrar intervalos de predicción no triviales, los cuales pueden dividirse en cuatro categorías: el método pivotal, el método de la prueba de inversión, el método plug-in (método naive) y los métodos que utilizan densidades predictivas aproximadas. Sólo los dos primeros métodos nos permiten encontrar intervalos de predicción exactos, mientras que los otros son métodos usados para encontrar intervalos de predicción aproximados. Debido a las aplicaciones realizadas en este trabajo, se describe a continuación la metodología plug-in o naive.

1.5.2.1. Método plug-in o naive

Ahora, sea $q_{\mathbf{Y}|\mathbf{X}}(\alpha; \theta)$ el cuantil α de $\mathbf{Y}|\mathbf{X} = \mathbf{X}$. Entonces, cualquier intervalo $[q_{\mathbf{Y}|\mathbf{X}}(\alpha_1; \theta), q_{\mathbf{Y}|\mathbf{X}}(1 - \alpha_2; \theta)]$, donde $\alpha_1 + \alpha_2 = \alpha$, tiene una probabilidad de $(1 - \alpha)$ de contener a \mathbf{Y} . Para obtener un intervalo de predicción para \mathbf{Y} , un método naive frecuentemente aplicado usa los valores procedentes de $\mathbf{X} = (X_1, \dots, X_n)$ para obtener un estimador de θ y sustituirlo en $q_{\mathbf{Y}|x}(\alpha; \theta)$. Este intervalo $[q_{\mathbf{Y}|\mathbf{X}}(\alpha_1; \hat{\theta}(\mathbf{X})), q_{\mathbf{Y}|\mathbf{X}}(1 - \alpha_2; \hat{\theta}(\mathbf{X}))]$ es llamado con frecuencia, intervalo de predicción plug-in. Este intervalo no es un intervalo de predicción $(1 - \alpha)$ exacto, pero condicionado a que $\hat{\theta}(\mathbf{X})$ tenga buenas propiedades asintóticas, este puede ser un intervalo de predicción $(1 - \alpha)$ aproximado cuando n es suficientemente grande. Por ejemplo, cuando $\hat{\theta}(\mathbf{X})$ es el estimador de máxima verosimilitud (MLE), se sabe que $[q_{\mathbf{Y}|\mathbf{X}}(\alpha_1; \hat{\theta}(\mathbf{X})), q_{\mathbf{Y}|\mathbf{X}}(1 - \alpha_2; \hat{\theta}(\mathbf{X}))]$ tenderá a $[q_{\mathbf{Y}|\mathbf{X}}(\alpha_1; \theta), q_{\mathbf{Y}|x}(1 - \alpha_2; \theta)]$ a medida que n aumente dado que el MLE converge casi seguramente a θ bajo algunas condiciones de regularidad. Note que la mayoría de las veces, los MLE son estimadores usados para obtener intervalos de predicción plug-in.

Los intervalos de predicción plug-in son muy estrechos para la mayoría de los problemas, ya que ignoran la incertidumbre en $\hat{\theta}(\mathbf{X})$ con respecto a θ . Por lo tanto, estos intervalos usualmente tienen una probabilidad de cobertura inferior a $(1 - \alpha)$. Este problema es más relevante para valores pequeños o moderados de n , a medida que n aumente, este problema eventualmente desaparecerá. Esta deficiencia de los intervalos plug-in puede ser parcialmente resuelta calibrando dichos intervalos.

1.5.3. Distribuciones predictivas

Una distribución predictiva es una distribución independiente de θ construida a partir de los datos \mathbf{X} observados con el fin de hacer declaraciones predictivas acerca del vector aleatoria \mathbf{Y} . Su función de distribución acumulada (cdf) será denotada por $\tilde{F}_p(y|x)$ y su densidad por $\tilde{f}_p(y|x)$. Sin pérdida de generalidad, se considera ahora el intervalo de predicción unilateral de la forma $(-\infty, L(\mathbf{X})]$ y simplemente denotado por $L_{1-\alpha}(\mathbf{X})$ un intervalo de predicción α exacto (o aproximado). Se puede utilizar la densidad predictiva para obtener tal intervalo al encontrar el cuantil $(1 - \alpha)$ de esta distribución, es decir

$$\tilde{F}_p(L_{1-\alpha}(x)|x) = 1 - \alpha.$$

Por lo tanto, graficando $\tilde{F}_p(y|x)$ a lo largo del rango de $\mathbf{Y}|x$ se obtendrán (exacta o aproximadamente) los intervalos de predicción $(1 - \alpha)$ para todo α ($0 < \alpha < 1$).

Por ejemplo, la densidad predictiva

$$\tilde{f}_p(y|x) = f(y|x; \hat{\theta}(x)),$$

puede ser usada para encontrar intervalos de predicción plug-in. Por otro lado, cuando $Z(\mathbf{X}, \mathbf{Y})$ tiene una distribución que es independiente de θ , el método pivotal tiene una cdf predictiva

$$\tilde{F}_p(y|x) = P[Z(\mathbf{X}, \mathbf{Y}) \leq Z(x, y)].$$

Cuando se busca un intervalo de predicción, no es fácil comparar los diferentes métodos de predicción para determinar cuál puede ser el adecuado. La única sugerencia que se ha hecho hasta ahora ha sido seleccionar entre una clase de métodos que proporcionen exactamente la misma probabilidad de cobertura, los métodos que den los intervalos más cortos (Fredette (2004)). Aunque, incluso cuando tal idea puede ser aplicada, se pueden obtener resultados diferentes para cualquier θ y para cualquier α . No obstante, dado que la gran mayoría de métodos predictivos pueden ser expresados a través de una distribución predictiva, se pueden comparar usando estas distribuciones. Esta comparación se puede hacer evaluando qué tan cerca está cada densidad de su densidad objetivo $f(y|x; \theta)$, para lo cual existen diferentes metodologías propuestas en la literatura, las cuales, de acuerdo con el objetivo de este trabajo, no resultan ser de interés. Sin embargo, sí se definirá un procedimiento conocido como calibración, el cual usualmente ayuda a “mover” una distribución predictiva cerca de su distribución objetivo. Este procedimiento es usado frecuentemente junto con el método plug-in o naive.

1.5.4. Calibración

Una familia de intervalos de predicción $L_{1-\alpha}(\mathbf{X})$ indexados por $(1 - \alpha)$ es obtenida a partir de una cdf predictiva tal que $\tilde{F}_p(L_{1-\alpha}(x)|x) = 1 - \alpha$ para cualquier α . Por lo tanto, $(1 - \alpha)$ no necesariamente representa la probabilidad de cobertura del intervalo de predicción: nos gustaría obtener intervalos con una probabilidad de cobertura de $(1 - \alpha)$, pero muchas de las distribuciones predictivas no satisfacen esto. Sea $H(1 - \alpha; \theta)$ la probabilidad de cobertura actual de $L_{1-\alpha}(x)$, es decir

$$P[\mathbf{Y} \leq L_{1-\alpha}(\mathbf{X}); \theta] = H(1 - \alpha; \theta).$$

Cuando los intervalos de predicción están asociados a un método de predicción exacto (frecuentista), $H(1 - \alpha; \theta) = 1 - \alpha$ para cualquier α y θ , pero $H(1 - \alpha; \theta)$ usualmente converge a $1 - \alpha$ a medida que la información acerca de θ contenida en \mathbf{X} aumente. En la práctica, mientras θ sea desconocido, $H(1 - \alpha; \theta)$ es usualmente aproximado por una

función $\check{H}(1 - \alpha)$ de α solamente. Este procedimiento es llamado calibración y proporciona una probabilidad de cobertura aproximada de $\check{H}(1 - \alpha)$ para el intervalo de predicción $(-\infty, L_{1-\alpha}(x)]$.

La calibración puede hacerse usando expansiones asintóticas o simulación. Ésta es usada frecuentemente en conjunto con el método plug-in. Por ejemplo Cox (1975) y Barndorff-Nielsen y Cox (1996) consideran el método plug-in y demuestran que bajo algunas condiciones de regularidad la probabilidad de cobertura puede ser escrita como

$$H(1 - \alpha; \theta) = 1 - \alpha + \frac{d(\theta)}{n} + O(n^{-3/2}), \quad (1.16)$$

donde $n = \dim(\mathbf{X})$. La idea es encontrar un nuevo intervalo, con frecuencia en la misma familia de intervalos de predicción (digamos $L_{1-\alpha'}(x)$) que absorba (o reduzca) el término $\frac{d(\theta)}{n}$ en 1.16. Vidoni (1995) y Komaki (1996) también utilizan expansiones asintóticas para mejorar el método plug-in.

La idea de aproximar $H(1 - \alpha; \theta)$ vía simulación es usualmente más tratable, de esto pueden encontrarse propuestas en Beran (1990) y en Meeker y Escobar (1998). La idea general del algoritmo presentado por Meeker y Escobar (1998) basado en bootstrap paramétrico, es obtener un valor $(1 - \alpha')$ tal que la probabilidad de cobertura aproximada $\check{H}(1 - \alpha')$ sea igual a $(1 - \alpha)$. A continuación se presenta el algoritmo original con una modificación que considera además el caso donde \mathbf{X} e \mathbf{Y} no son independientes:

1. Escoger un valor de $1 - \alpha$, digamos $1 - \alpha_0$ (preferiblemente mayor que $1 - \alpha$).
2. Simular $\mathbf{X}_1^*, \dots, \mathbf{X}_B^*$ a partir de $f(\mathbf{X}^*; \hat{\theta}(\mathbf{X}))$ donde $\hat{\theta}(\mathbf{X})$ es el MLE basado en la muestra \mathbf{X} observada.
3. Simular $\mathbf{Y}_1^*, \dots, \mathbf{Y}_B^*$ a partir de las densidades condicionales $f(\mathbf{Y}|\mathbf{X}^*; \hat{\theta}(\mathbf{X}))$.
4. Obtener los MLE $\hat{\theta}(\mathbf{X}_i^*)$ donde $i = 1, \dots, B$.
5. Calcular los B intervalos de predicción $1 - \alpha_0$ con el método plug-in basados en $(\mathbf{X}_i^*, \hat{\theta}(\mathbf{X}_i^*))$.
6. Aproximar la función de probabilidad de cobertura en $1 - \alpha_0$ por:

$$\check{H}(1 - \alpha_0) = \frac{1}{B} \sum_{i=1}^B \mathbb{I}(\mathbf{Y}_i^* \leq L_{1-\alpha_0}(\mathbf{X}_i^*)),$$

Donde $\mathbb{I}(A)$ es la variable indicadora del evento A .

7. Repetir los pasos 2-6 para diferentes valores de $1 - \alpha_0$ hasta que un $1 - \alpha'$ tal que $\check{H}(1 - \alpha') \approx 1 - \alpha$ sea encontrado.

Este algoritmo nos proporciona un intervalo de predicción con una probabilidad de cobertura aproximada de $(1 - \alpha)$.

1.6. Predicción estadística basada en datos de vida censurados

Las definiciones presentadas en esta sección son tomadas de Meeker y Escobar (1998) y Escobar y Meeker (1999). En esta sección se describen métodos que usan datos de vida censurados para construir límites o intervalos de predicción para resultados futuros, particularmente para datos cuya distribución pertenezca a la familia de log-localización-escala (por ejemplo, weibull o lognormal). Se considera específicamente el problema de predicción para una nueva muestra. Por lo tanto, se ilustra la metodología para construir límites o intervalos de predicción para los tiempos de falla futuros de una nueva unidad, usando datos de una muestra previa. El método general usa calibración basada en simulación.

En general, para predecir futuras realizaciones de una cantidad aleatorias se necesita lo siguiente:

1. Un modelo estadístico que describa la población o proceso de interés. Este modelo usualmente consiste en una distribución que depende de un vector de parámetros θ .
2. Información sobre los valores de los parámetros θ . Esta información puede venir de una prueba de vida de laboratorio o de datos de campo.

Asumiremos que los tiempos de falla siguen una distribución continua con cdf $F(t) = F(t; \theta)$ y fdp $f(t) = f(t; \theta)$, donde θ es un vector de parámetros. Generalmente, θ es desconocido y puede ser estimado a partir de la muestra de datos disponibles. En tales casos, asumiremos lo siguiente: independencia estadística de los tiempos de falla y que los tiempos de censura son independientes de cualquier tiempo de falla futuro que pueda ser observado si una unidad no queda censurada.

Se consideran situaciones en las cuales n unidades comienzan a operar en el tiempo 0 y son observadas hasta el tiempo t_c cuando los datos no están disponibles para ser analizados. Los tiempos de falla son registrados para las r unidades que fallan en el intervalo $(0, t_c]$. Por lo tanto, los datos consisten en los r estadísticos de orden más pequeños $T_{(1)} < \dots < T_{(r)} \leq T_c$ y la información de que las otras $n - r$ unidades habrán fallado después de t_c . Con datos con censura de tiempo (o Tipo I), t_c es preespecificado y r es aleatorio. Con datos con censura por falla (o Tipo II), r es preespecificado y $t_c = t_r$ es aleatorio.

1.6.1. Conceptos sobre los intervalos de predicción

1.6.1.1. Intervalos de predicción para tiempos de falla (dado θ)

Asumiendo que los tiempos de falla tienen una distribución de probabilidad continua completamente especificada $F(t; \theta)$ y una fdp $f(t; \theta)$, un intervalo de probabilidad de predicción exacto al $(1 - \alpha)100\%$ para una observación futura T de $F(t; \theta)$, donde θ es un vector de parámetros está dado por:

$$\text{PI}(1 - \alpha) = [T, \tilde{T}] = [t_{\alpha/2}, t_{1-\alpha/2}], \quad (1.17)$$

donde t_p es el cuantil p de $F(t; \theta)$. Por la definición de cuantiles de distribuciones continuas, la probabilidad de cobertura de el intervalo en (1.17) está dada por:

$$\begin{aligned}
\Pr[T \in \text{PI}(1 - \alpha); \theta] &= \Pr(\underline{T} \leq T \leq \tilde{T}; \theta) \\
&= \Pr(t_{\alpha/2} \leq T \leq t_{1-\alpha/2}; \theta) \\
&= 1 - \alpha.
\end{aligned} \tag{1.18}$$

1.6.1.2. Probabilidad de cobertura para procedimientos de intervalos de predicción estadística (θ estimado)

En predicción estadística de tiempos de vida, el objetivo es predecir una cantidad aleatoria T basándose en información muestral (denotada por DATOS). Generalmente, con base sólo en datos muestrales, existe incertidumbre en la distribución de los parámetros. Los DATOS aleatorios llevan a un parámetro estimado $\hat{\theta}$ y luego a un intervalo de predicción nominal $(1 - \alpha)100\% \text{ PI}(1 - \alpha) = [\underline{T}, \tilde{T}]$. Por lo tanto, $[\underline{T}, \tilde{T}]$ y la futura variable aleatoria T tienen una distribución conjunta que depende del vector de parámetros θ .

Existen dos tipos de Probabilidades de Cobertura (CP):

1. Para DATOS fijos (y por tanto $\hat{\theta}$ y $[\underline{T}, \tilde{T}]$ fijos) la probabilidad de cobertura condicional de un intervalo particular $[\underline{T}, \tilde{T}]$ es:

$$\begin{aligned}
\text{CP}(\text{PI}(1 - \alpha)|\hat{\theta}; \theta) &= \Pr(\underline{T} \leq T \leq \tilde{T}|\hat{\theta}; \theta) \\
&= F(\tilde{T}; \theta) - F(\underline{T}; \theta).
\end{aligned} \tag{1.19}$$

Esta probabilidad condicional es desconocida porque $F(t; \theta)$ depende de θ que es desconocido.

2. De muestra a muestra, la probabilidad de cobertura condicional es aleatoria porque $F(t; \theta)$ depende de $\hat{\theta}$. La probabilidad de cobertura incondicional para el procedimiento de intervalos de predicción es:

$$\begin{aligned}
\text{CP}(\text{PI}(1 - \alpha); \theta) &= \Pr(\underline{T} \leq T \leq \tilde{T}; \theta) \\
&= E_{\hat{\theta}} \left\{ \text{CP}(\text{PI}(1 - \alpha)|\hat{\theta}; \theta) \right\},
\end{aligned} \tag{1.20}$$

donde el valor esperado es con respecto a $\hat{\theta}$ aleatorio. Debido a que (1.20) puede ser calculada (al menos aproximadamente) y puede ser controlada, esta probabilidad incondicional es la que generalmente es utilizada para describir un procedimiento de intervalo de predicción.

Cuando $\text{CP}(\text{PI}(1 - \alpha); \theta) = 1 - \alpha$ no depende de θ , se dice que el procedimiento PI (por sus siglas en inglés) es “exacto”. Cuando $\text{CP}(\text{PI}(1 - \alpha); \theta) \neq 1 - \alpha$ no depende de θ , generalmente es posible encontrar un procedimiento PI modificado que sea “exacto”. Cuando $\text{CP}(\text{PI}(1 - \alpha); \theta)$ depende de un θ desconocido, se dice que PI es un procedimiento de intervalo de predicción aproximado. En estos casos puede ser posible modificar un procedimiento específico para encontrar una mejor aproximación.

1.6.1.3. Relación entre límites de predicción unilaterales e intervalos de predicción bilaterales

Un límite de predicción unilateral inferior al $(1 - \alpha/2)100\%$ para una observación futura de $F(t; \hat{\theta})$ está dado por

$$\text{PI}(1 - \alpha/2) = [\underline{T}, \infty) = [\hat{t}_\alpha, \infty) = [t_\alpha(\hat{\theta}), \infty)$$

donde

$$\Pr(\underline{T} \leq T < \infty) = 1 - \alpha/2.$$

Un límite de predicción unilateral superior al $(1 - \alpha/2)100\%$ para una observación futura de $F(t; \hat{\theta})$ está dado por

$$\text{PI}(1 - \alpha/2) = (0, \tilde{T}] = (0, \hat{t}_{1-\alpha}] = (0, t_{1-\alpha}(\hat{\theta})]$$

donde

$$\Pr(0 < T \leq \tilde{T}) = 1 - \alpha/2.$$

Combinando un límite de predicción unilateral inferior $(1 - \alpha/2)100\%$ y un límite de predicción unilateral superior $(1 - \alpha/2)100\%$ se obtiene un intervalo de predicción $(1 - \alpha)100\%$ de igual cola. En particular, si $\Pr(\underline{T} \leq T < \infty) = 1 - \alpha/2$ y $\Pr(0 < T \leq \tilde{T}) = 1 - \alpha/2$, entonces $\Pr(\underline{T} \leq T \leq \tilde{T}) = 1 - \alpha$. Puede ser posible encontrar un intervalo estrecho con probabilidades diferentes en las colas superior e inferior, que sigan sumando α . El uso de intervalos de predicción de igual cola, sin embargo, tiene la importante ventaja de proveer un intervalo que tiene puntos finales que pueden ser interpretados correctamente como límites de predicción unilaterales (con el ajuste apropiado en el nivel de confianza). Cuando se calculan intervalos de predicción bilaterales, frecuentemente es necesario calcular por separado los límites de predicción inferior y superior y luego unirlos para obtener dicho intervalo de predicción.

1.6.1.4. El método naive (plug-in) para calcular intervalos de predicción estadística

Un intervalo de predicción naive para T continua se obtiene al sustituir el estimador de máxima verosimilitud (MLE) para θ en (1.17), dando

$$\text{PI}(1 - \alpha) = [\underline{T}, \tilde{T}] = [\hat{t}_{\alpha/2}, \hat{t}_{1-\alpha/2}],$$

donde $\hat{t}_p = t_p(\hat{\theta})$ es el MLE para el cuantil p de T . Por ejemplo, para predecir una observación futura independiente de una distribución de log-localización-escala con cdf $\Pr[T \leq t] = \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$, un intervalo de predicción naive está dado por:

$$\begin{aligned}
\text{PI}(1 - \alpha) &= [\underline{T}, \tilde{T}] = [\hat{t}_{\alpha/2}, \hat{t}_{1-\alpha/2}] \\
&= [\exp(\hat{\mu} + \Phi^{-1}(\alpha/2) \times \hat{\sigma}), \\
&\quad \exp(\hat{\mu} + \Phi^{-1}(1 - \alpha/2) \times \hat{\sigma})].
\end{aligned}$$

La probabilidad de cobertura incondicional para este procedimiento naive es aproximadamente igual al $(1 - \alpha)$ nominal con tamaños de muestra grandes. Sin embargo, para un número moderado o pequeño de unidades fallando, la probabilidad de cobertura puede no ser cercana a $(1 - \alpha)$.

1.6.2. Calibración de límites de predicción estadística naive (plug-in)

Cox (1975) sugiere un método aproximado de muestra grande, basado en los MLE, que puede ser usado para calibrar o corregir un intervalo de predicción naive. Atwood (1984) usa un método similar. La idea básica de este enfoque, es calibrar el límite de predicción unilateral naive evaluando la función $\text{CP}[\text{PI}(1 - \alpha_c); \theta]$ en $\hat{\theta}$ y encontrar un valor de calibración $(1 - \alpha_c)$, tal que, para un límite inferior de predicción unilateral para T

$$\begin{aligned}
\text{CP}[\text{PI}(1 - \alpha_{cl}); \hat{\theta}] &= \Pr(\underline{T} \leq T \leq \infty; \hat{\theta}) \\
&= \Pr(\hat{t}_{\alpha_{cl}} \leq T \leq \infty; \hat{\theta}) = 1 - \alpha
\end{aligned} \tag{1.21}$$

y para un límite de predicción unilateral superior para T

$$\begin{aligned}
\text{CP}[\text{PI}(1 - \alpha_{cu}); \hat{\theta}] &= \Pr(0 \leq T \leq \tilde{T}; \hat{\theta}) \\
&= \Pr(0 \leq T \leq \hat{t}_{1-\alpha_{cu}}; \hat{\theta}) = 1 - \alpha.
\end{aligned} \tag{1.22}$$

En problemas en los cuales $\text{CP}[\text{PI}(1 - \alpha_c); \theta]$ no depende de θ , el procedimiento de calibración proporciona un intervalo de predicción exacto. En algunos casos simples, la curva de calibración puede obtenerse analíticamente en términos de los cuantiles de distribuciones estándar.

1.6.2.1. Calibración aproximada de los límites de predicción estadística naive

Cox (1975) sugiere una aproximación analítica asintótica para (1.21) y (1.22). Para calibrar un límite de predicción inferior naive (el método es similar para el límite de predicción superior), sea $\text{PI}(1 - \alpha) = [\underline{T}, \infty] = [\hat{t}_{\alpha}, \infty] = [t_{\alpha}(\hat{\theta}), \infty]$. La probabilidad de cobertura condicional de $\text{PI}(1 - \alpha)$ es una función de $(1 - \alpha)$, $\hat{\theta}$ y θ , digamos

$$\begin{aligned}
\text{CP}[\text{PI}(1 - \alpha)|\hat{\theta}; \theta] &= \Pr(\underline{T} \leq T < \infty|\hat{\theta}; \theta) \\
&= g(\alpha, \hat{\theta}; \theta),
\end{aligned}$$

de donde, la probabilidad de cubrimiento incondicional de $\text{PI}(1 - \alpha)$ es $\text{CP}[\text{PI}(1 - \alpha); \theta] = E_{\hat{\theta}}[g(\alpha, \hat{\theta}; \theta)]$. Bajo condiciones de regularidad estándar, tomando el valor esperado de la expansión de las series de Taylor de $g(\alpha, \hat{\theta}; \theta)$ da

$$\begin{aligned} \text{CP}[\text{PI}(1 - \alpha); \theta] &= 1 - \alpha + \frac{1}{n} \sum_{i=1}^k a_i \left. \frac{\partial g(\alpha, \hat{\theta}; \theta)}{\partial \hat{\theta}_i} \right|_{\theta} \\ &\quad + \frac{1}{2n} \sum_{i,j=1}^k b_{ij} \left. \frac{\partial^2 g(\alpha, \hat{\theta}; \theta)}{\partial \hat{\theta}_i \partial \hat{\theta}_j} \right|_{\theta} + o\left(\frac{1}{n}\right). \end{aligned}$$

donde a_i y b_{ij} son elementos del vector \mathbf{a} y la matriz \mathbf{B} , respectivamente, tal como se define por

$$E_{\hat{\theta}}[(\hat{\theta} - \theta)] = \mathbf{a}(\theta) + o\left(\frac{1}{n}\right)$$

$$E_{\hat{\theta}}[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] = \mathbf{B}(\theta) + o\left(\frac{1}{n}\right).$$

Incluso para problemas aparentemente simples, estos valores esperados son extremadamente difíciles de calcular. En las pocas situaciones en las cuales la expresión es tratable (por ejemplo, muestras sin censura de distribuciones exponenciales y normales), ya existen procedimientos de predicción exactos simples basados en la distribución de cantidades pivotaes.

1.6.2.2. Método 1: Calibración por simulación del proceso de muestreo/ predicción

Las capacidades computacionales modernas facilitan el uso de los métodos de Monte Carlo para evaluar, numéricamente, cantidades como (1.21) y (1.22), incluso para modelos estadísticos complicados. Beran (1990) proporciona teoría asintótica para dicha calibración de la predicción. En particular, para un límite de predicción unilateral inferior, bajo ciertas condiciones de regularidad y sin censura, Beran demuestra que la probabilidad de cobertura incondicional para un procedimiento de predicción calibrado $\text{PI}(1 - \alpha_{cl})$ es

$$\text{CP}[\text{PI}(1 - \alpha_{cl}); \theta] = 1 - \alpha + O\left(\frac{1}{n^2}\right)$$

En otras palabras, la dependencia de la CP en θ rápidamente disminuye a medida que n aumenta. Este resultado se mantiene para límites de predicción unilaterales superiores y para dos límites unilaterales usados juntos para formar un intervalo de predicción bilateral. Beran (1990) además muestra que el orden de la aproximación asintótica puede mejorarse iterando el procedimiento de calibración pero indicando que la realización de la aproximación de orden superior puede no ser tan buena en muestras pequeñas.

Para calibrar vía simulación bajo el modelo asumido, podemos usar los MLE $\hat{\theta}$ para simular tanto el muestreo como el proceso de predicción un número grande de veces B . Aunque $B = 2000$ se sugiere con frecuencia para intervalos de confianza basados en simulación, generalmente son requeridos valores de B más grandes para problemas de predicción, debido a la variabilidad añadida a la futura observación.

Conceptualmente, para obtener las curvas de calibración para los límites de predicción inferior y superior, la función $\text{CP}[\text{PI}(1 - \alpha_c); \hat{\theta}]$ en (1.21) y (1.22) puede ser evaluada como lo indica el siguiente algoritmo:

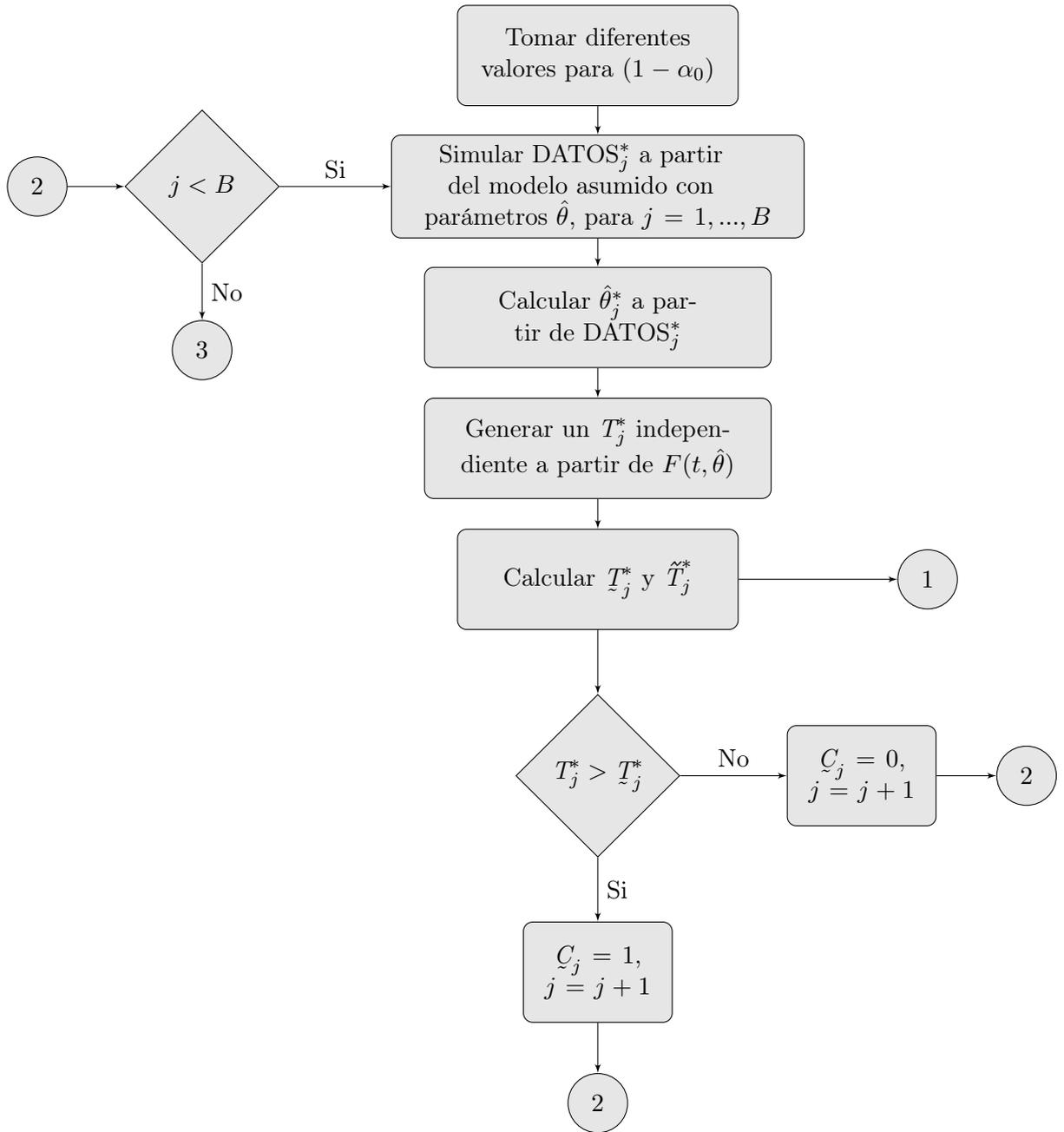


Figura 1.2: Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 1)

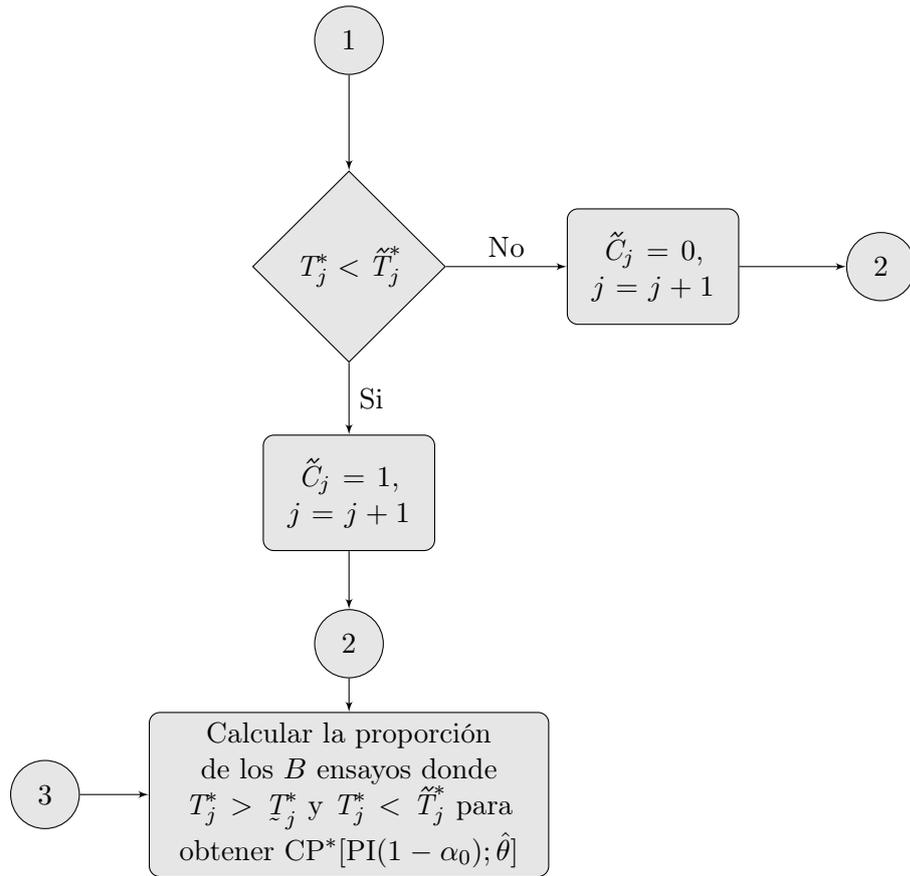


Figura 1.3: Continuación Figura 1.2: Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 1)

La diferencia entre $CP[PI(1-\alpha_0); \hat{\theta}]$ y $CP^*[PI(1-\alpha_0); \hat{\theta}]$ es debida el error de Monte Carlo y puede hacerse pequeña arbitrariamente escogiendo un valor de B lo suficientemente grande.

Método 2: Operacionalmente, la función $CP[PI(1-\alpha_c); \hat{\theta}]$ completa en (1.21) y (1.22) puede ser evaluada más directamente, pero equivalentemente, usando el procedimiento descrito en el siguiente algoritmo:

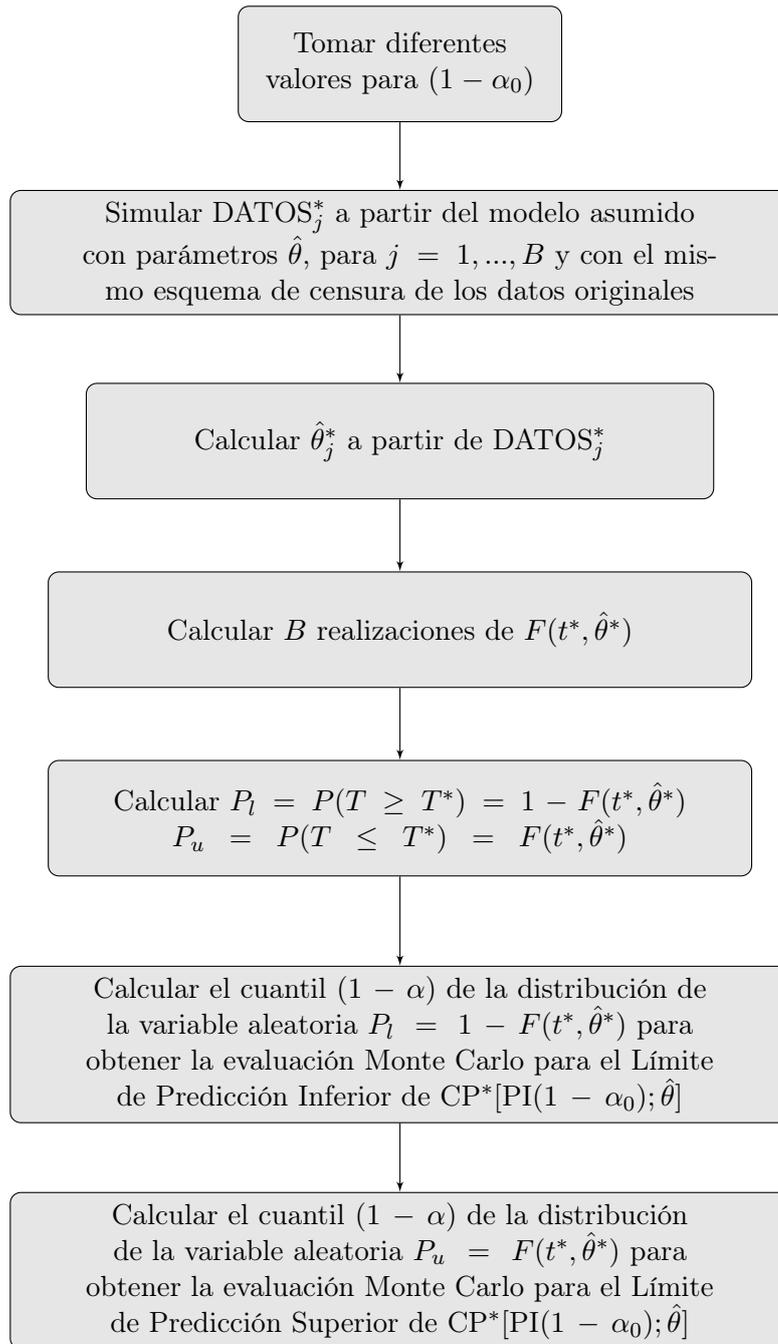


Figura 1.4: Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 2)

Para ejemplificar esta metodología, a continuación se presenta la equivalencia entre el procedimiento de calibración y un procedimiento pivotal (aproximadamente pivotal) para intervalos de predicción a partir de distribuciones de log-localización-escala, el cual podría ser utilizado si en nuestro problema se estuvieran considerando solo los datos arrojados por la prueba de laboratorio o ALT.

1.6.2.3. Equivalencia entre el procedimiento de calibración y el procedimiento pivotal (aproximadamente pivotal) para intervalos de predicción a partir de distribuciones de log-localización-escala

Con censura Tipo II (falla), una prueba de vida se ejecuta hasta que ocurra un número r de fallas especificado, donde $1 \leq r \leq n$. Para datos completos o con censura Tipo II, $Z_{\log(T)} = [\log(T) - \hat{\mu}]/\hat{\sigma}$ es pivotal. Esto significa que, la distribución de $Z_{\log(T)}$ depende solo de n y r y no de μ y σ . Para censura de tiempo (la prueba de vida se ejecuta hasta un tiempo de censura t_c especificado), $Z_{\log(T)}$ es solo aproximadamente pivotal y los cuantiles de $Z_{\log(T)}$ dependen de $F(t_c; \mu, \sigma)$ (la proporción de falla esperada desconocida en el tiempo t_c) y el tamaño de muestra n (ver Antle y Rademaker (1972) y Engelhardt y Bain (1979)).

En general, para un límite de predicción inferior en un problema de predicción para una distribución de log-localización-escala

$$\begin{aligned} \text{CP}[\text{PI}(1 - \alpha_{cl}); \mu, \sigma] &= \Pr[T \geq \hat{t}_{\alpha_{cl}}; \mu, \sigma] \\ &= \Pr[\log(T) \geq \hat{\mu} + \hat{\sigma}\Phi^{-1}(\alpha_{cl}); \mu, \sigma] \\ &= \Pr[\log(T) - \hat{\mu}]/\hat{\sigma} \geq \Phi^{-1}(\alpha_{cl}); \mu, \sigma] \\ &= 1 - \Pr[Z_{\log(T)} \leq \Phi^{-1}(\alpha_{cl})] = 1 - \alpha. \end{aligned} \quad (1.23)$$

Consecuentemente, $\Pr[Z_{\log(T)} \leq \Phi^{-1}(\alpha_{cl})] = \alpha$ o equivalentemente, el cuantil α de la distribución de $Z_{\log(T)}$ es $\Phi^{-1}(\alpha_{cl})$; es decir $z_{\log(T)(\alpha)} = \Phi^{-1}(\alpha_{cl})$. Esto lleva a que $1 - \alpha_{cl} = 1 - \Phi(Z_{\log(T)(\alpha)})$ es el cuantil $1 - \alpha$ de la distribución de $P = [1 - \Phi(Z_{\log(T)(\alpha)})]$.

Cuando $Z_{\log(T)}$ es pivotal, la probabilidad de cobertura en (1.23) no depende de (μ, σ) . Por lo tanto, el límite de predicción inferior pivotal es $\hat{\mu} + \hat{\sigma} \times z_{\log(T)(\alpha)}$ y el límite de predicción inferior calibrado (aproximadamente calibrado) es $\hat{\mu} + \hat{\sigma} \times \Phi^{-1}(\alpha_{cl})$. Señalando que $z_{\log(T)(\alpha)} = \Phi^{-1}(\alpha_{cl})$ muestra que los dos procedimientos de predicción dan el mismo límite de predicción.

Cuando $Z_{\log(T)}$ no es pivotal, (1.23) se evalúa en los MLE $\mu = \hat{\mu}$ y $\sigma = \hat{\sigma}$, dando una calibración aproximada. Esta evaluación se expresa como

$$\begin{aligned} \text{CP}[\text{PI}(1 - \alpha_{cl}); \hat{\mu}, \hat{\sigma}] &= \Pr[\log(T^*) - \hat{\mu}^*/\hat{\sigma}^* \geq \Phi^{-1}(\alpha_{cl}); \hat{\mu}^*, \hat{\sigma}^*] \\ &= 1 - \Pr[Z_{\log(T^*)} \leq \Phi^{-1}(\alpha_{cl})] = 1 - \alpha. \end{aligned}$$

El límite de predicción inferior aproximadamente pivotal es $\hat{\mu} + \hat{\sigma} \times z_{\log(T^*)(\alpha)}$ y el límite de predicción inferior calibrado (aproximadamente calibrado) es $\hat{\mu} + \hat{\sigma} \times \Phi^{-1}(\alpha_{cl})$. Señalando que $z_{\log(T^*)(\alpha)} = \Phi^{-1}(\alpha_{cl})$ muestra que los dos procedimientos de predicción dan el mismo límite de predicción.

Cuanto los cuantiles de la distribución de $Z_{\log(T)}$ (o $Z_{\log(T^*)}$) no están disponibles, usamos simulación para obtener (o aproximar) los cuantiles de la distribución de estas dos variables aleatorias. Para una simulación de tamaño B , el procedimiento pivotal (aproximadamente pivotal) usa el cuantil α , $z_{\log(T^*)(\alpha)}$, de la distribución empírica de los valores simulados

$Z_{\log(T_j^*)} = [\log(T_j^*) - \hat{\mu}_j^*] / \hat{\sigma}_j^*$, $j = 1, \dots, B$, para construir el límite de predicción inferior, $\hat{\mu} + \hat{\sigma} \times z_{\log(T^*)(\alpha)}$.

Además, la distribución empírica de los valores observados de $P^* = [1 - \Phi(Z_{\log(T^*)})]$ proporciona una evaluación de $\text{CP}^*[\text{PI}(1 - \alpha_{cl}); \hat{\mu}, \hat{\sigma}]$. Entonces, a partir de los resultados inmediatamente siguientes (1.23), $1 - \alpha_{cl} = [1 - \Phi(z_{\log(T^*)(\alpha)})]$ y el límite de predicción inferior calibrado (aproximadamente calibrado) es $\hat{\mu} + \hat{\sigma} \times \Phi^{-1}(\alpha_{cl})$. De nuevo, $z_{\log(T^*)(\alpha)} = \Phi^{-1}(\alpha_{cl})$, mostrando que los dos procedimientos dan límites de predicción inferiores idénticos.

Cuando los datos están completos o tienen censura Tipo II, la única diferencia entre los cuantiles $z_{\log(T)(\alpha)}$ y $z_{\log(T^*)(\alpha)}$ es debida al error de Monte Carlo, y la cobertura de los intervalos de predicción puede hacerse tan cercana a $1 - \alpha$ como se quiera tomando un valor grande para el tamaño de simulaciones B .

Para el límite de predicción superior, la calibración consiste en encontrar el $1 - \alpha_{cu}$ tal que

$$\begin{aligned} \text{CP}[\text{PI}(1 - \alpha_{cu}); \mu, \sigma] &= \Pr[T \leq \hat{t}_{1 - \alpha_{cu}}; \mu, \sigma] \\ &= \Pr[\log(T) \leq \hat{\mu} + \hat{\sigma} \Phi^{-1}(1 - \alpha_{cu}); \mu, \sigma] \\ &= \Pr[Z_{\log(T)} \leq \Phi^{-1}(1 - \alpha_{cu}); \mu, \sigma] = 1 - \alpha. \end{aligned}$$

Por lo tanto, el cuantil $1 - \alpha$ de la distribución de $Z_{\log(T)}$ es $\Phi^{-1}(1 - \alpha_{cu})$. Esto lleva a que $1 - \alpha_{cu} = \Phi(z_{\log(T)(1 - \alpha)})$ es el cuantil $1 - \alpha$ de la distribución de $P = [\Phi(Z_{\log(T)})]$. Cuando los cuantiles de la distribución de $Z_{\log(T)}$ no están disponibles, usamos simulación nuevamente para aproximar los cuantiles y los límites de predicción se obtienen reemplazando $z_{\log(T)(1 - \alpha)}$ por $z_{\log(T^*)(1 - \alpha)}$ y $1 - \alpha_{cu} = \Phi(z_{\log(T)(1 - \alpha)})$ por $1 - \alpha_{cu} = \Phi(z_{\log(T^*)(1 - \alpha)})$.

El límite de predicción superior pivotal (o aproximadamente pivotal) es $\hat{\mu} + \hat{\sigma} \times z_{\log(T^*)(\alpha)}$ y el límite de predicción superior calibrado es $\hat{\mu} + \hat{\sigma} \times \Phi^{-1}(1 - \alpha_{cu})$. Señalando que $z_{\log(T^*)(\alpha)} = \Phi^{-1}(1 - \alpha_{cu})$, esto muestra que los dos procedimientos de predicción dan el mismo límite de predicción superior.

1.6.2.4. Método 3: Calibración promediando las probabilidades de cobertura condicional

Puede ser mucho más eficiente, computacionalmente, obtener las curvas de calibración en (1.21) y (1.22) simulando las probabilidades de cobertura condicional como las dadas en (1.19) y promediarlas para estimar el valor esperado en 1.20 (ver Mee y Kushary (1994)). El procedimiento a seguir se describe en el siguiente algoritmo:

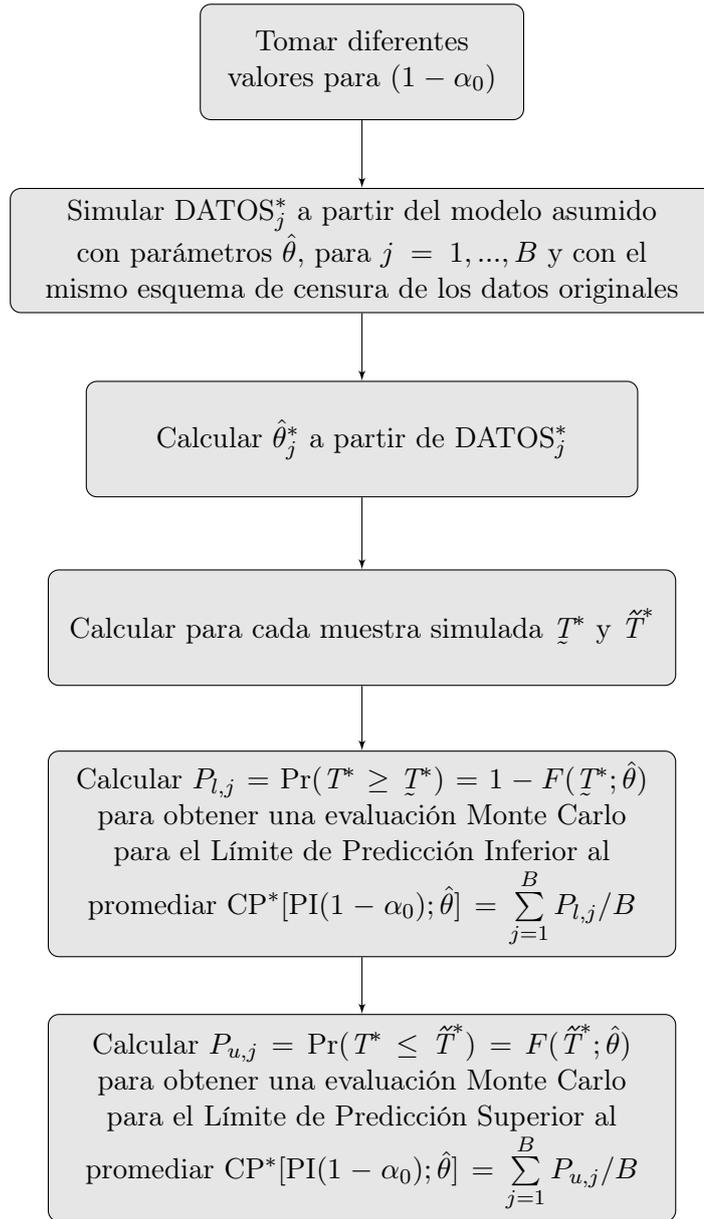


Figura 1.5: Algoritmo para la calibración de los límites de predicción inferior y superior promediando las probabilidades de cobertura condicional (Método 3)

Operacionalmente, para calcular un límite de predicción unilateral inferior sólo se necesita encontrar el valor $1 - \alpha_{cl}$ apropiado. $CP^*[PI(1 - \alpha_c); \hat{\theta}]$ es una función continua y creciente de $1 - \alpha_c$, por lo tanto, el valor apropiado de calibración puede ser encontrado usando un método de hallazgo de raíz.

El procedimiento para la evaluación Monte Carlo de la probabilidad de cobertura que se muestra en 1.6.2.2 usa la proporción observada de los límites de predicción correctos. La ventaja del procedimiento promediando la probabilidad, es que este no requiere una simulación de la variable aleatoria futura en la evaluación. Entonces, el procedimiento

necesita menos muestras Monte Carlo para obtener el mismo nivel de precisión. El método descrito en 1.6.2.2 podría ser escogido en situaciones en las cuales el intervalo de predicción naive es fácil de calcular pero las probabilidades condicionales no pueden ser calculadas fácilmente (por ejemplo, cuando la cdf y los cuantiles de la variable aleatoria a ser predecida no pueden ser calculadas en forma cerrada).

Predicción estadística para el modelo de tasa de uso

2.1. Un problema con el componente A

El problema con en el componente A que se describe detalladamente en la Sección 1.4 y con base en el cual Meeker *et al.* (2009) plantean el modelo de tasa de uso para confiabilidad en campo, es real, sin embargo, debido a que la accesibilidad a este tipo de información es limitada, en este trabajo de investigación, como ya se mencionó, se generarán los datos para el aparato o electrodoméstico en cuestión vía simulación usando el paquete estadístico R (2011), teniendo en cuenta sus características específicas de uso, todo esto con el fin de obtener los intervalos de predicción para el tiempo de vida de este producto una vez se encuentre en manos del cliente.

2.1.1. Simulación de los datos para el problema con el componente A

El primer conjunto de datos que es necesario obtener es el de los ciclos de uso promedio por semana (\mathbf{R}) del aparato en los hogares y sus probabilidades asociadas ($\boldsymbol{\pi}$), los cuales varían entre los mismos y se distribuyen de acuerdo a una distribución lognormal discretizada truncada en 20 usos, esto debido a que, según Meeker *et al.* (2009), a partir de la encuesta realizada por el departamento de marketing de la compañía se concluyó que el número de hogares con más de 20 usos por semana era insignificante. Por lo tanto, para simular la distribución estimada de los ciclos de uso por semana, se generó en R una distribución lognormal discretizada con media $\mu = 1.6$ y desviación estándar $\sigma = 0.5$. En el **apéndice A** se presenta el código implementado en R con el cual se construyó dicha distribución y que se muestra en la Figura 1.1.

En la Tabla 2.1 se presentan los valores para \mathbf{R} y $\boldsymbol{\pi}$ obtenidos a partir de la simulación anterior

Tabla 2.1: Tasas de uso simuladas del componente A en los hogares y sus probabilidades correspondientes.

R_i	π_i	R_i	π_i
1	0.0083314953	11	0.0212949409
2	0.0772192256	12	0.0133945580
3	0.1564536504	13	0.0090734094
4	0.1832527922	14	0.0066872531
5	0.1618074633	15	0.0045918470
6	0.1224058070	16	0.0028974755
7	0.0898016884	17	0.0019750957
8	0.0631028052	18	0.0018848629
9	0.0431312787	19	0.0010126125
10	0.0308295402	20	0.0008521987

Luego de obtener los valores para las tasas de uso del producto en los hogares (campo), el siguiente paso es simular las muestras que corresponden a los ciclos hasta la falla en la ALT y a los tiempos de vida del producto en campo cuya distribución es el modelo propuesto por Meeker *et al.* (2009), esto con el fin de estimar, a partir de dichas muestras, los parámetros η_C y σ_C .

En la Sección 1.4 se presenta la distribución para los ciclos, C , hasta la falla dada por la ALT como una distribución del tipo log-localización-escala. En este caso, debido a que los datos son simulados, se asume que la distribución para los ciclos hasta la falla determinada en el laboratorio es lognormal. Por lo tanto, dicha distribución es de la forma

$$F_C(c) = \Pr(C \leq c) = \Phi_{nor} \left[\frac{\log(c/\eta_C)}{\sigma_C} \right] = \Phi_{nor} \left[\frac{\log(c) - \log(\eta_C)}{\sigma_C} \right], \quad (2.1)$$

donde Φ es la cdf normal, η_C es el parámetro de escala, la mediana en este caso, dado en unidades de los ciclos de la prueba y σ_C es el parámetro de forma.

Entonces,

$$\log(C) \sim \mathbf{Nor}(\log(\eta_C), \sigma_C^2). \quad (2.2)$$

Bajo esta distribución, se generó en R una muestra de tamaño $m = 200$ con valores $\eta_C = 200$ y $\sigma_C = 1$ (ver **apéndice A**); estos valores se presentan en la Tabla 2.2.

Tabla 2.2: Ciclos hasta la falla en la prueba de laboratorio (ALT) (Datos simulados).

Ciclos hasta la falla en la ALT (m=200)									
153.79	268.33	50.95	85.65	85.44	150.49	103.83	242.75	524.57	271.82
98.12	300.35	33.93	115.60	72.13	75.38	18.71	71.58	113.51	181.56
164.18	107.49	179.25	110.60	798.26	569.78	741.54	66.81	1446.10	203.01
36.35	129.24	85.51	129.72	436.48	282.73	188.88	102.10	144.72	99.31
122.01	218.71	189.56	122.43	218.94	246.10	70.67	381.70	423.77	96.62
125.77	460.55	441.26	204.38	574.57	487.54	235.79	399.57	1001.59	66.43
60.18	388.26	527.04	1039.95	210.54	33.62	593.16	199.45	49.21	70.39
419.26	25.70	103.70	113.42	177.78	66.62	60.40	177.16	201.48	406.33
157.12	121.93	206.02	75.01	199.92	218.17	772.33	121.86	2726.95	56.17
4955.30	1405.98	1048.73	46.52	1307.19	117.73	87.76	166.88	237.01	361.12
224.42	77.85	90.39	386.62	482.99	160.88	86.23	45.54	77.77	607.00
467.21	509.83	424.84	168.49	853.19	380.83	70.96	62.78	72.78	37.33
567.13	39.14	218.55	101.19	66.02	72.63	154.04	279.50	311.49	1464.00
490.47	99.77	579.29	541.41	1096.28	444.53	181.77	205.94	454.87	150.76
423.79	747.66	389.20	537.05	49.96	1041.82	728.35	504.61	147.63	191.65
41.28	235.05	81.63	81.94	72.79	160.51	116.57	816.70	21.59	255.58
334.14	555.99	95.88	260.73	1815.02	371.51	171.65	157.08	137.72	422.89
65.83	308.90	141.67	88.70	352.11	49.87	161.90	150.18	584.57	179.32
52.12	63.31	155.19	102.40	189.88	297.09	192.08	895.20	474.84	458.22
194.02	22.82	283.91	402.87	269.89	195.28	226.48	474.31	655.28	1265.19

Ahora, ya se determinó que la distribución de la tasa de uso promedio \mathbf{R} (ciclos de uso por semana) en manos del cliente está definida por una distribución discreta con valores:

$$\mathbf{R} = (R_1, R_2, \dots, R_k), \text{ con } k = 1, 2, \dots, 20,$$

con probabilidades correspondientes:

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k), \text{ donde } \sum_{i=1}^k \pi_i = 1.$$

Como se especificó en la Sección 1.4, si los ciclos de prueba en la ALT tienen el mismo efecto sobre el tiempo de vida que los ciclos de uso del producto en servicio, entonces la fracción de falla en el tiempo real t en la población del producto (semanas de servicio), está descrita por una distribución de mezcla finita, que para este caso, debido a que los datos son simulados, se asume que es la siguiente mezcla finita de lognormales:

$$F(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \Pr(T \leq t) = \sum_{i=1}^k \pi_i \Phi_{nor} \left[\frac{\log(R_i \times t / \eta_C)}{\sigma_C} \right].$$

Resolviendo el logaritmo, puede reescribirse de la forma

$$F(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \Pr(T \leq t) = \sum_{i=1}^k \pi_i \Phi_{nor} \left[\frac{\log(t) - \log(\eta_C / R_i)}{\sigma_C} \right]. \quad (2.3)$$

Por lo tanto,

$$(\log(T)|\mathbf{R} = R_i) \sim \mathbf{Nor}(\log(\eta_C/R_i), \sigma_C^2). \quad (2.4)$$

Para obtener la muestra correspondiente a el tiempo de vida del producto en campo (semanas de servicio), de acuerdo a el modelo en (2.3), primero es necesario obtener una muestra para las tasas de uso \mathbf{R} a partir de la distribución que previamente fue determinada para dicha variable (ver Figura 1.1 y Tabla 2.1). Este es un caso típico de generación de variables aleatorias discretas, por lo tanto, se utilizó el paquete estadístico R para generar $n = 100$ valores correspondientes a las tasas de uso del componente A en los hogares (ver Morgan (1984)). Una vez obtenidos estos datos, fue posible determinar una muestra del mismo tamaño para los tiempos de vida del producto en manos del cliente, teniendo en cuenta su distribución dada en (2.4), para esto se tomaron nuevamente los valores de $\eta_C = 200$ y $\sigma_C = 1$. En el **apéndice A** se presenta el código implementado y en la Tabla 2.3 se pueden ver los valores obtenidos.

Tabla 2.3: Tiempo de vida del producto en campo (Datos simulados).

Tiempo de vida en semanas del producto en manos del cliente (n=100)									
20.98	27.69	9.95	132.31	50.19	20.47	22.31	22.64	212.99	3.47
104.42	91.29	22.48	22.84	238.63	20.12	163.23	7.20	9.55	145.80
64.05	16.38	14.65	253.83	77.73	41.93	25.09	20.68	45.63	19.34
149.90	234.73	51.31	48.22	58.57	28.95	6.30	27.38	55.92	55.61
90.75	54.16	100.39	17.45	29.70	36.57	40.26	82.70	19.90	24.41
21.48	42.61	153.69	40.52	100.85	45.22	4.29	68.54	5.28	69.42
39.79	45.16	225.29	37.92	33.76	21.83	26.39	18.13	35.79	40.31
47.63	20.26	84.82	76.91	61.55	87.84	81.65	25.29	17.51	7.61
366.66	6.14	8.37	51.07	79.40	99.36	28.40	88.31	171.37	27.09
29.78	67.48	60.36	24.13	207.65	15.13	8.73	11.58	65.05	5.33

2.1.2. Estimación por máxima verosimilitud para el modelo de tasa de uso para el componente A

Una vez simulados los datos de las tasas de uso del componente A en campo y las muestras correspondientes a los 200 ciclos hasta la falla obtenidos en el laboratorio y a los 100 tiempos de vida del producto en manos del cliente, el paso a seguir y que puede considerarse como uno de los principales en el proceso de predicción estadística, es el de la estimación de los parámetros del modelo. En este caso se sabe de ante mano que la distribución de los ciclos hasta la falla en la ALT es lognormal y que la de los tiempos de vida en campo es una mezcla finita de lognormales. Otro aspecto importante que se va a considerar, es el hecho de que la estimación se va a realizar para datos de vida sin censura y con censura. A continuación se presenta el procedimiento implementado en el paquete estadístico R para la estimación de los parámetros η_C y σ_C para los dos tipos de datos anteriormente mencionados, así como los resultados obtenidos.

2.1.2.1. Estimación por máxima verosimilitud para el modelo de tasa de uso para el componente A con datos de vida sin censura

En la Sección 1.4 se presentaron las funciones de verosimilitud y log verosimilitud genéricas para la ALT y para campo. Como se determinó que la distribución de los ciclos hasta la falla en la ALT es lognormal y que la de los tiempos de vida en campo es una mezcla finita de lognormales, dichas funciones para datos de vida sin censura son más específicas y se muestran a continuación.

La función de verosimilitud para los datos sin censura obtenidos a partir de la ALT puede expresarse como

$$L_{ALT}(c; \eta_C, \sigma_C) = \prod_{j=1}^m \frac{1}{c_j \sigma_C} \phi_{nor} \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right].$$

donde ϕ es la fdp normal, η_C es el parámetro de escala, la mediana en este caso, dado en unidades de los ciclos de la prueba y σ_C es el parámetro de forma.

Ahora, la función de log verosimilitud para los datos sin censura obtenidos a partir de la ALT es

$$\mathcal{L}_{ALT}(c; \eta_C, \sigma_C) = \sum_{j=1}^m \left\{ -\log(c_j) - \log(\sigma_C) + \log \left(\phi_{nor} \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right) \right\}.$$

Por otro lado, la función de verosimilitud para la mezcla finita que describe la fracción de falla en el tiempo real t para el producto en campo, en la escala de tiempo de servicio (semanas), para los datos completos sin censura puede ser escrita en la forma

$$L_{Campo}(t_{comp}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \prod_{j=1}^n \sum_{i=1}^k \pi_i^{z_{ij}} \left(\phi_{nor} \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}}.$$

donde cada valor de R_i representa la tasa de uso promedio del producto en manos del cliente en unidades de ciclos de uso por semana y cada π es su probabilidad correspondiente, además cada $\mathbf{z}_j = (z_{ij}, i = 1, \dots, k)$ es un vector indicador de magnitud k con 1 en la posición correspondiente a la categoría apropiada y 0 en caso contrario dentro de la mezcla, es decir, que indica si el individuo j proviene del componente i de la mezcla.

Luego, la función de log verosimilitud para los datos en campo sin censura es

$$\mathcal{L}_{Campo}(t_{comp}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \sum_{j=1}^n \log \left\{ \sum_{i=1}^k \pi_i^{z_{ij}} \left(\phi_{nor} \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}} \right\}.$$

Ahora, el log de la verosimilitud total para los datos de campo y los datos en la ALT sin censura está dada por

$$\mathcal{L}(t_{\text{comp}}, c; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) + \mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C). \quad (2.5)$$

Esta función de log verosimilitud total representa la probabilidad conjunta de los datos, asumiendo que las $n = 100$ (tiempo de vida en campo) y $m = 200$ (ciclos hasta la falla en la ALT) son observaciones independientes. Como los valores para $\boldsymbol{\pi}$ son conocidos, ya que provienen de la distribución de la tasa de uso del producto que fue estimada por el departamento de marketing de la compañía, los Z_{ij} también son conocidos, esto debido a que para cada observación se conoce la categoría de uso en el campo a la cual pertenece, por lo tanto es posible realizar el cálculo explícito de los MLE maximizando la función de log verosimilitud total en (2.5). Para esto se puede recurrir a métodos como el de Newton-Rapson, Nelder-Mead o al método conocido como Evolución Diferencial (ver Beyer y Schwefel (2002)), este último implementado en el paquete estadístico R en la librería **DEoptim**.

En la Tabla 2.4 se presenta la estimación para η_C y σ_C , en unidades de los ciclos hasta la falla en la prueba de laboratorio, para los datos de vida del componente A sin censura realizada por medio del método de máxima verosimilitud vía Evolución Diferencial utilizando la librería **DEoptim**. En el **apéndice A** se puede encontrar el código implementado en R para la obtención de éstos MLE.

Tabla 2.4: Parámetros estimados vía Evolución Diferencial para los datos de vida del componente A sin censura.

Parámetro	Estimación
$\hat{\eta}_C$	200.3825271
$\hat{\sigma}_C$	0.6803828

Al tener datos de vida no censurados en la ALT, como en este caso, se puede evidenciar que la prueba hecha a los productos en el laboratorio se realizó hasta que todas las unidades fallaron sin fijar un tiempo límite de observación. Para el caso del producto en campo, el hecho de tener datos de vida no censurados, quiere decir que se recibieron reclamos por garantías de todas las unidades que salieron al mismo tiempo al mercado, por lo tanto se conocen los tiempos exactos de falla para cada una de ellas.

2.1.2.2. Estimación por máxima verosimilitud para el modelo de tasa de uso para el componente A con datos de vida censurados

En la Sección 1.4, para el escenario con censura, se estableció que en la prueba de laboratorio o ALT se presenta una censura a derecha no aleatoria, esto debido a que la prueba se realiza hasta un número de ciclos hasta la falla predeterminado, por lo tanto el número de unidades que no hayan fallado hasta ese momento quedarán censuradas. Por este motivo, para realizar la estimación de los parámetros por máxima verosimilitud, primero es necesario definir el número de ciclos en la ALT hasta los cuales se hace la observación, es decir, los ciclos que determinarán la censura.

Para el problema con el componente A, se tomaron tres valores para el número de ciclos a partir de los cuales quedarían censuradas las unidades en prueba, estos valores corresponden a los percentiles 95, 80 y 70 de la distribución de los $m = 200$ ciclos hasta la falla lognormales generados en un principio. Dichos valores se presentan en la Tabla 2.5.

Tabla 2.5: Ciclos hasta la falla en la ALT que determinan la censura

Percentil	Ciclos en la ALT
95	1042
80	474
70	381

Entonces, una vez definidos los ciclos de censura, se puede establecer la función de verosimilitud para los datos censurados obtenidos a partir de la ALT y que lleva información de observaciones exactas (no censuradas) y de observaciones con censura a derecha no aleatoria. Dicha función está dada por

$$L_{\text{ALT}}(c; \eta_C, \sigma_C) = \prod_{j=1}^m \left\{ \left(\frac{1}{c_j \sigma_C} \phi_{\text{nor}} \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{\tau_j} \times \left(1 - \Phi_{\text{nor}} \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{1-\tau_j} \right\},$$

donde τ_j es el indicador de censura que toma valores de $\tau_j = 1$ si c_j es una observación exacta en la ALT y $\tau_j = 0$ si c_j es una observación censurada a derecha (ciclos hasta el fin de la prueba sin observar falla) en la ALT.

Luego, la función de log verosimilitud para los datos censurados obtenidos a partir de la ALT es

$$\begin{aligned} \mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C) = & \sum_{j=1}^m \tau_j \left\{ -\log(c_j) - \log(\sigma_C) + \log \left(\phi_{\text{nor}} \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right) \right\} \\ & + (1 - \tau_j) \left\{ \log \left(1 - \Phi_{\text{nor}} \left[\frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right) \right\}. \end{aligned}$$

Por otro lado, como ya se mencionó previamente, los datos de retornos por garantía son censurados cuando no todas las unidades que están en manos del cliente han fallado durante este periodo de tiempo. Es decir que en este segundo escenario, las unidades en campo tienen una censura a derecha no aleatoria. Por lo tanto, para realizar la estimación de los parámetros, se debe definir este tiempo de garantía ya que va a determinar las unidades que en campo quedarán censuradas. Según lo anterior, el tiempo de garantía para nuestro problema, coincide con el tiempo de censura (t_c).

Nuevamente, para el problema con el componente A, se tomaron tres valores para determinar el tiempo de censura t_c del producto en manos del cliente, estos valores corresponden

a los percentiles 95, 80 y 70 de la distribución de los $n = 100$ tiempos de vida en campo generados a partir de la mezcla de lognormales, es decir que se tomaron tres tiempos distintos como periodos de garantía.

Para calcular estos percentiles, se utilizaron las funciones **norMix** y **qnorMix** de la librería **nor1mix** del paquete estadístico R (ver Mächler (2011)). Esta librería se enfoca en los modelos de mezcla de normales univariadas, por lo tanto permite la evaluación de funciones de densidad y de funciones de densidad acumulada, el cálculo de cuantiles, la generación de valores aleatorios y la realización de gráficos para este tipo de modelos, entre otras posibilidades. Como el modelo de tasa de uso es una mezcla univariada de lognormales (ver (2.4)), para que esta librería se ajustara a nuestras necesidades, los resultados obtenidos a partir de sus funciones se exponenciaron para ser llevados a la escala original y así obtener una estimación aproximada de dichos percentiles. Este procedimiento puede ser realizado debido a que se dá la siguiente relación.

De (2.4) se deduce que

$$(T|\mathbf{R} = R_i) \sim \mathbf{Lognormal}(\eta_C/R_i, \sigma_C^2).$$

Por la definición de cuantiles t_p es tal que

$$P(T|\mathbf{R} = R_i \leq t_p) = p$$

donde p es la probabilidad correspondiente al cuantil t_p . Ahora,

$$P(\ln(T|\mathbf{R} = R_i) \leq \ln(t_p)) = p.$$

Sea $Y = \ln(T|\mathbf{R} = R_i)$ y $y^* = \ln(t_p)$, entonces

$$P(Y \leq y^*) = p$$

y^* es el cuantil p de Y , es decir y_p . Por lo tanto

$$y_p = \ln t_p$$

$$\exp(y_p) = t_p.$$

Al utilizar la función **norMix** lo que se hace es crear un objeto que representa la mezcla univariada de normales con todos sus parámetros, que en este caso, son los valores del $\log(\eta_C/R_i)$, σ_C y $\boldsymbol{\pi}$ (ver (2.4)), para después emplear la función **qnorMix** que permite el cálculo de los percentiles a partir del objeto de clase **norMix** creado en un principio, posteriormente, por ser una mezcla de lognormales, dichos percentiles fueron llevados a la escala original, por lo cual se sabe que existe la presencia de sesgo en su estimación (ver **apéndice A**).

Debido a que por definición la función cuantil está dada por

$$F^{-1}(p) = \inf \{t_p \in \mathbb{R} : F(t_p) \geq p\},$$

donde p es la probabilidad correspondiente al cuantil t_p , la función **qnorMix**, internamente lo que hace es resolver el problema de encontrar la raíz de la ecuación

$$F(t_p; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) - p \geq 0,$$

es decir,

$$\sum_{i=1}^k \pi_i \Phi_{nor} \left[\frac{\log(t_p) - \log(\eta_C/R_i)}{\sigma_C} \right] - p \geq 0,$$

también es posible utilizar la función **uniroot** del paquete estadístico R para resolver el problema de encontrar los cuantiles (percentiles) de la mezcla requeridos. Los resultados obtenidos al utilizar ambos métodos para encontrar los percentiles son muy similares, por lo que se llegó a la conclusión de que el sesgo presente al implementar la función **qnorMix** puede ser ignorado y es por eso que este método fue el escogido para hacer las aproximaciones.

Los percentiles encontrados, que corresponden a los tiempos de censura, t_c , en unidades de las semanas y años de servicio del producto en campo, se presentan en la Tabla 2.6.

Tabla 2.6: Tiempos de censura en campo

Percentil	Tiempo de censura en semanas (t_c)	Tiempo de censura en años (t_c)
95	257	5.4
80	104	2.2
70	73	1.5

Luego, la función de verosimilitud para la mezcla finita que describe la fracción de falla en el tiempo real t para el producto en campo, en la escala de tiempo de servicio (semanas), para los datos completos censurados en el tiempo t_c , puede ser escrita en la forma

$$L_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \prod_{j=1}^n \left\{ \sum_{i=1}^k \pi_i^{z_{ij}} \left(\phi_{nor} \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}} \right\}^{\delta_j} \\ \times \left\{ 1 - \sum_{i=1}^k \pi_i^{z_{ij}} \left(\Phi_{nor} \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{z_{ij}} \right\}^{1-\delta_j},$$

donde δ_j es el indicador de censura que toma valores de $\delta_j = 1$ si t_j es una observación exacta en campo y $\delta_j = 0$ si t_j es una observación censurada a derecha (tiempos de falla no observados) en campo.

Luego, la función de log verosimilitud para los datos en campo censurados es

$$\begin{aligned} \mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) &= \sum_{j=1}^n \delta_j \log \left\{ \sum_{i=1}^k \pi_i^{z_{ij}} \left(\phi_{\text{nor}} \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}} \right\} \\ &+ (1 - \delta_j) \log \left\{ 1 - \sum_{i=1}^k \pi_i^{z_{ij}} \left(\Phi_{\text{nor}} \left[\frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{z_{ij}} \right\}. \end{aligned}$$

Ahora, la log verosimilitud total para los datos de campo y los datos en la ALT censurados está dada por

$$\mathcal{L}(t_{\text{comp}}, c; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) + \mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C). \quad (2.6)$$

Como se mencionó en la estimación para datos de vida sin censura, los valores de los $\boldsymbol{\pi}$ y los Z_{ij} son conocidos, ya que provienen de la distribución de las tasas de uso del producto en campo \mathbf{R} , la cual fue estimada por el departamento de marketing de la compañía a partir de la información obtenida en un estudio previo, por lo tanto, es posible realizar el cálculo explícito de los MLE maximizando la función de log verosimilitud total en (2.6).

En la Tabla 2.7 se presenta la estimación para η_C y σ_C , en unidades de los ciclos hasta la falla en la prueba de laboratorio, para los datos de vida censurados a derecha del componente A, realizada por medio del método de máxima verosimilitud vía Evolución Diferencial utilizando la librería **DEoptim**. En el **apéndice A** se puede encontrar el código implementado en R para la obtención de éstos MLE.

Para facilitar la notación con respecto a los tiempos de censura, denotaremos como **Censura 1** al tiempo de censura a derecha dado por el percentil 95 tanto en los ciclos hasta la falla en la ALT como en los tiempos de vida en el campo, como **Censura 2** al tiempo de censura a derecha dado por el percentil 80 y como **Censura 3** al tiempo de censura a derecha dado por el percentil 70 (ver Tabla 2.5 y Tabla 2.6).

Tabla 2.7: Parámetros estimados vía Evolución Diferencial para los datos de vida del componente A censurados.

Tiempo de censura a derecha	Parámetro	Estimación
Censura 1 (Percentil 95)	$\hat{\eta}_C$	198.7205186
	$\hat{\sigma}_C$	0.6721745
Censura 2 (Percentil 80)	$\hat{\eta}_C$	192.1097325
	$\hat{\sigma}_C$	0.6594054
Censura 3 (Percentil 70)	$\hat{\eta}_C$	190.0611496
	$\hat{\sigma}_C$	0.6683192

Como se puede observar, a medida que el tiempo de censura disminuye y por lo tanto se tiene menos tiempo para observar los ciclos hasta la falla de las unidades en la ALT y los tiempos de garantía para el producto en manos del cliente son menores, el parámetro $\hat{\eta}_C$ toma un valor menor, lo cual indica que tiende a subestimarse. Este resultado se da por el hecho de que en el momento de realizar la estimación se cuenta con menos información acerca de los tiempos de falla exactos de las unidades tanto en la ALT como en el campo. Por otro lado, la estimación dada por $\hat{\sigma}_C$, no presenta mucha variación entre los diferentes tiempos de censura.

2.1.3. Predicción estadística para el modelo de tasa de uso para el componente A

Después del procedimiento de estimación de los parámetros del modelo de tasa de uso tanto para datos de vida sin censura como censurados, el paso a seguir con el fin de cumplir los objetivos de este trabajo de investigación, es el cálculo de los intervalos de predicción bilaterales y los límites de predicción unilaterales plug-in (naive) para estos dos tipos de datos de vida, los cuales ya se han descrito ampliamente en secciones anteriores.

Como se mencionó en la Sección 1.6, un intervalo de predicción bilateral naive para T al $(1 - \alpha)100\%$ está dado por

$$\text{PI}(1 - \alpha) = [\underline{T}, \tilde{T}] = [\hat{t}_{\alpha/2}, \hat{t}_{1-\alpha/2}],$$

donde, $\hat{t}_p = t_p(\hat{\eta}_C, \hat{\sigma}_C)$ es el MLE para el cuantil p de T .

Análogamente, el límite de predicción unilateral inferior naive al $(1 - \alpha)100\%$ puede expresarse de la forma

$$\text{PI}(1 - \alpha_l) = [\underline{T}, +\infty) = [\hat{t}_\alpha, \infty).$$

Además, el límite de predicción unilateral superior naive al $(1 - \alpha)100\%$ puede expresarse de la forma

$$\text{PI}(1 - \alpha_u) = (0, \tilde{T}] = (0, \hat{t}_{1-\alpha}].$$

Para calcular estos intervalos y límites de predicción para el modelo de tasa de uso propuesto por Meeker *et al.* (2009) dado en (2.3), es necesario determinar los cuantiles para dicha distribución, según la probabilidad de cubrimiento deseada. Debido a que este modelo está descrito por una mezcla de lognormales, se utilizaron las funciones **normMix** y **qnorMix** de la librería **nor1mix** del paquete estadístico R para aproximar estos valores. La programación puede verse en el **apéndice A**.

A continuación se presentan los intervalos de predicción bilaterales $\text{PI}(1 - \alpha)$ naive y los límites de predicción unilaterales inferiores $\text{PI}(1 - \alpha_l)$ y superiores $\text{PI}(1 - \alpha_u)$ naive con probabilidades de cobertura correspondientes $\text{CP}[\text{PI}(1 - \alpha); \hat{\theta}]$, $\text{CP}[\text{PI}(1 - \alpha_l); \hat{\theta}]$ y $\text{CP}[\text{PI}(1 - \alpha_u); \hat{\theta}]$ iguales a 0.90 y 0.95, todo esto para el tiempo de vida del componente A en campo en unidades de las semanas de servicio con los diferentes tiempos de censura:

Tabla 2.8: Intervalos de predicción bilaterales naive sin censura

Sin Censura	
CP[PI(1 - α); $\hat{\theta}$]	PI(1 - α)
0.90	[10.19789, 164.47993]
0.95	[7.845625, 215.916508]

Tabla 2.9: Límites de predicción unilaterales inferiores naive sin censura

Sin Censura	
CP[PI(1 - α_l); $\hat{\theta}$]	PI(1 - α_l)
0.90	[13.8146, $+\infty$)
0.95	[10.19789, $+\infty$)

Tabla 2.10: Límites de predicción unilaterales superiores naive sin censura

Sin Censura	
CP[PI(1 - α_u); $\hat{\theta}$]	PI(1 - α_u)
0.90	(0, 120.4639]
0.95	(0, 164.47993]

Tabla 2.11: Intervalos de predicción bilaterales naive con Censura 1

Con Censura 1	
CP[PI(1 - α); $\hat{\theta}$]	PI(1 - α)
0.90	[10.22435, 161.36658]
0.95	[7.882787, 211.408194]

Tabla 2.12: Límites de predicción unilaterales inferiores naive con Censura 1

Con Censura 1	
CP[PI(1 - α_l); $\hat{\theta}$]	PI(1 - α_l)
0.90	[13.81653, + ∞)
0.95	[10.22435, + ∞)

Tabla 2.13: Límites de predicción unilaterales superiores naive con Censura 1

Con Censura 1	
CP[PI(1 - α_u); $\hat{\theta}$]	PI(1 - α_u)
0.90	(0, 118.45863]
0.95	(0, 161.36658]

Tabla 2.14: Intervalos de predicción bilaterales naive con Censura 2

Con Censura 2	
CP[PI(1 - α); $\hat{\theta}$]	PI(1 - α)
0.90	[10.05261, 153.41521]
0.95	[7.77604, 200.37783]

Tabla 2.15: Límites de predicción unilaterales inferiores naive con Censura 2

Con Censura 2	
CP[PI(1 - α_l); $\hat{\theta}$]	PI(1 - α_l)
0.90	[13.53302, + ∞)
0.95	[10.05261, + ∞)

Tabla 2.16: Límites de predicción unilaterales superiores naive con Censura 2

Con Censura 2	
CP[PI(1 - α_u); $\hat{\theta}$]	PI(1 - α_u)
0.90	(0, 113.02918]
0.95	(0, 153.41521]

Tabla 2.17: Intervalos de predicción bilaterales naive con Censura 3

Con Censura 3	
CP[PI(1 - α); $\hat{\theta}$]	PI(1 - α)
0.90	[9.828942, 153.556945]
0.95	[7.585515, 200.990628]

Tabla 2.18: Límites de predicción unilaterales inferiores naive con Censura 3

Con Censura 3	
CP[PI(1 - α_l); $\hat{\theta}$]	PI(1 - α_l)
0.90	[13.26696, + ∞)
0.95	[9.828942, + ∞)

Tabla 2.19: Límites de predicción unilaterales superiores naive con Censura 3

Con Censura 3	
CP[PI(1 - α_u); $\hat{\theta}$]	PI(1 - α_u)
0.90	(0, 112.84895]
0.95	(0, 153.556945]

Como puede verse en las Tablas anteriores y como se mencionó en la Sección 1.6, combinando un límite de predicción unilateral inferior $(1 - \alpha/2)100\%$ y un límite de predicción unilateral superior $(1 - \alpha/2)100\%$ se obtiene un intervalo de predicción $(1 - \alpha)100\%$ de igual cola. Es decir, que si $\Pr(\underline{T} \leq T < \infty) = 1 - \alpha/2$ y $\Pr(0 < T \leq \check{T}) = 1 - \alpha/2$, entonces $\Pr(\underline{T} \leq T \leq \check{T}) = 1 - \alpha$. Por lo tanto, para calcular un intervalo de predicción bilateral para el tiempo de vida del producto en campo al 90%, los límites de predicción unilaterales naive al 95% pueden ser combinados.

Por otro lado, los intervalos y límites de predicción naive (plug-in) son generalmente muy estrechos y su probabilidad de cobertura es inferior al valor nominal $1 - \alpha$ debido a que ignoran la incertidumbre, en este caso, de $\hat{\eta}_C$ y $\hat{\sigma}_C$ con respecto a η_C y σ_C , es por esto que el procedimiento no es exacto sino aproximado. Para resolver esta deficiencia, es posible recurrir a los métodos de calibración propuestos por Escobar y Meeker (1999) para mejorar de alguna manera la aproximación obtenida a partir del procedimiento naive.

En las siguientes Subsecciones se presentan los resultados obtenidos al aplicar las técnicas de calibración para los límites de predicción unilaterales naive para el modelo de tasa uso para confiabilidad en campo para el componente A, las cuales son propuestas por Escobar y Meeker (1999) y se describen ampliamente en la Sección 1.6 de este trabajo.

2.1.3.1. Calibración de los límites de predicción unilaterales naive para el modelo de tasa de uso para el componente A

Para obtener las curvas de calibración que se presentan a continuación, se implementaron en el programa estadístico R el **Método 1**, descrito en el algoritmo de la Figura 1.2 y la Figura 1.3 y el **Método 3**, descrito en el algoritmo de la Figura 1.5, ambos con un tamaño de muestra en la simulación de $B = 5.000$. El **Método 2**, que se describe en la Figura 1.4, se trabajó con un tamaño de muestra en la simulación de $B = 50.000$.

En la ejecución de estos algoritmos aplicados al modelo de tasa de uso para el componente A, se reemplazó la función $F(t, \hat{\theta})$ por la función en (2.3) evaluada en los parámetros estimados de acuerdo al esquema de censura establecido, así

$$F(t; \hat{\eta}_C, \hat{\sigma}_C, \boldsymbol{\pi}, \mathbf{R}) = \Pr(T \leq t) = \sum_{i=1}^k \pi_i \Phi_{nor} \left[\frac{\log(t) - \log(\hat{\eta}_C/R_i)}{\hat{\sigma}_C} \right].$$

Por otro lado, para la graficación y con el fin de disminuir el número de simulaciones necesarias, los resultados fueron suavizados a partir de la técnica de interpolación spline cúbica utilizando la función **smooth.spline** del programa estadístico R (ver Durbán (2009)). Además, las probabilidades de cobertura que se tomaron para la construcción de los intervalos y límites de predicción calibrados fueron 0.90 y 0.95.

Finalmente, para calcular los límites de predicción unilaterales sólo se necesitó encontrar el valor $1 - \alpha_c$ apropiado. Debido a que $\text{CP}^*[\text{PI}(1 - \alpha_c); \hat{\theta}]$ es una función continua y creciente de $1 - \alpha_c$, el valor apropiado de calibración se encontró usando el método de hallazgo de raíz que proporciona la función **uniroot** de R. La programación puede verse en el **apéndice A**.

- Curvas de calibración para datos de vida sin censura:

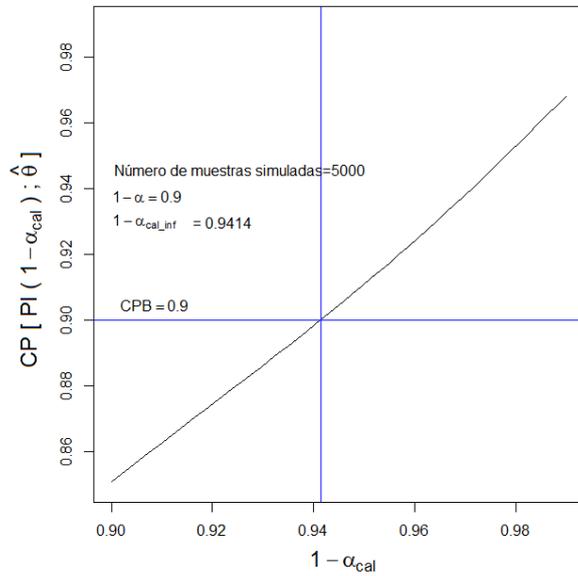


Figura 2.1: Curva de calibración del límite inferior por el Método 1 sin censura (CP=0.90)

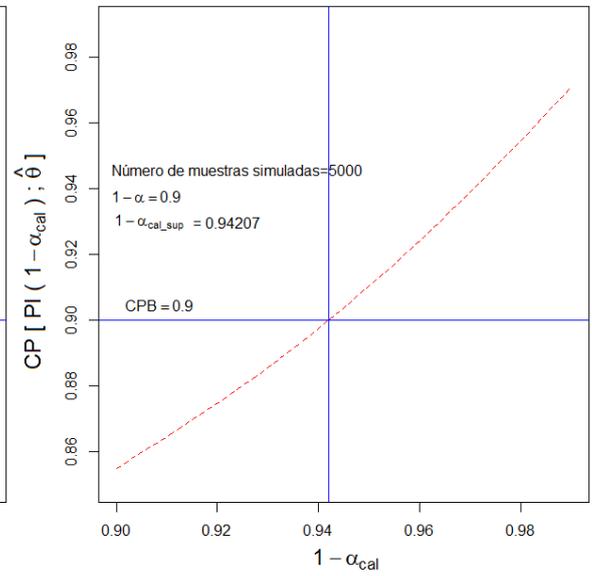


Figura 2.2: Curva de calibración del límite superior por el Método 1 sin censura (CP=0.90)

Usando los puntos de calibración con el Método 1 sin censura en la Figura 2.1 y en la Figura 2.2, se puede afirmar que un límite de predicción unilateral inferior naive al 94.14% proporciona un límite de predicción unilateral inferior calibrado aproximado al 90% para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.21% proporciona un límite de predicción unilateral superior calibrado aproximado al 90% para el tiempo de vida del producto en campo.

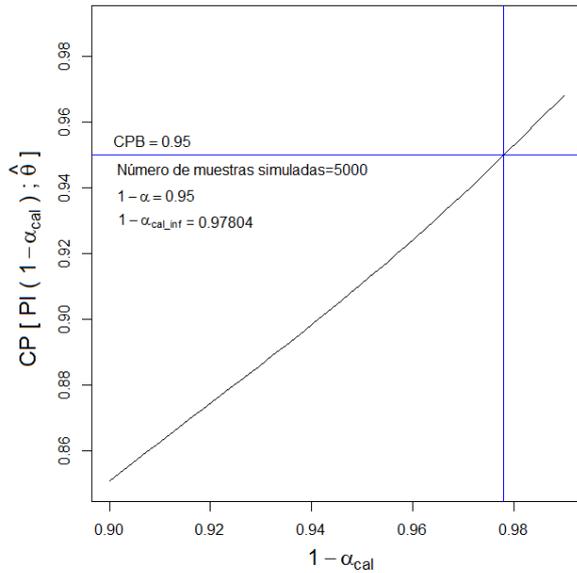


Figura 2.3: Curva de calibración del límite inferior por el Método 1 sin censura (CP=0.95)

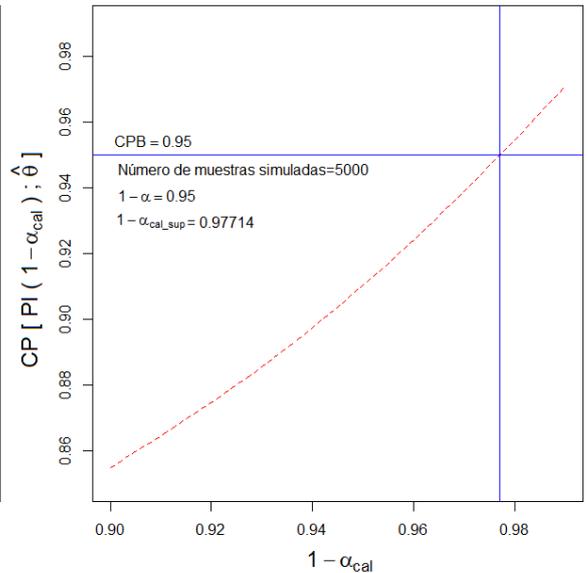


Figura 2.4: Curva de calibración del límite superior por el Método 1 sin censura (CP=0.95)

Usando los puntos de calibración con el Método 1 sin censura en la Figura 2.3 y en la Figura 2.4, se puede afirmar que un límite de predicción unilateral inferior naive al 97.80 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.71 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

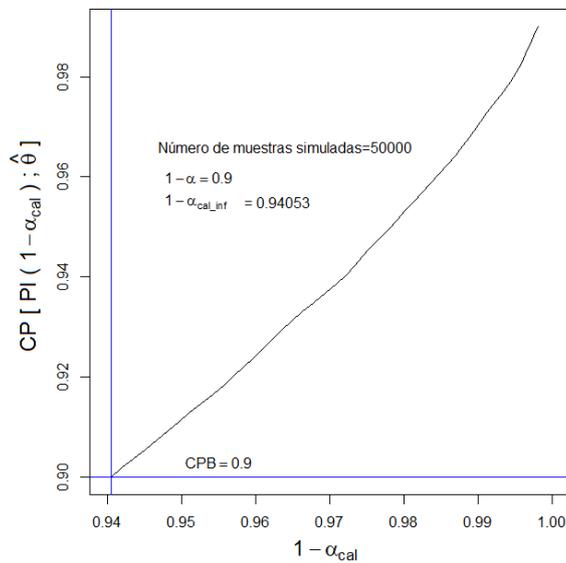


Figura 2.5: Curva de calibración del límite inferior por el Método 2 sin censura (CP=0.90)

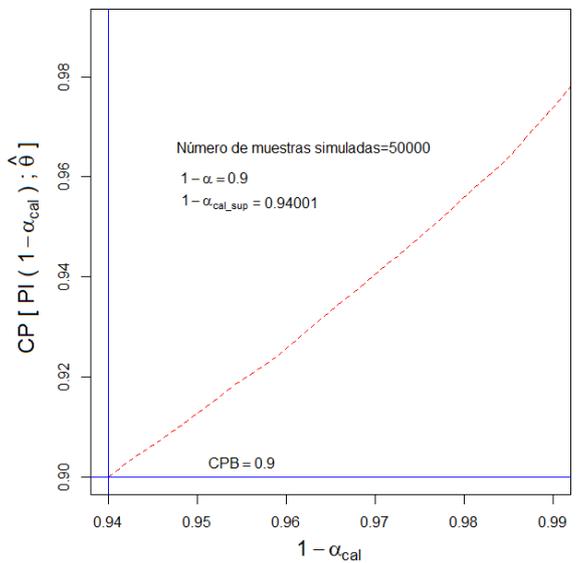


Figura 2.6: Curva de calibración del límite superior por el Método 2 sin censura (CP=0.90)

Usando los puntos de calibración con el Método 2 sin censura en la Figura 2.5 y en la Figura 2.6, se puede afirmar que un límite de predicción unilateral inferior naive al 94.05 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.00 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

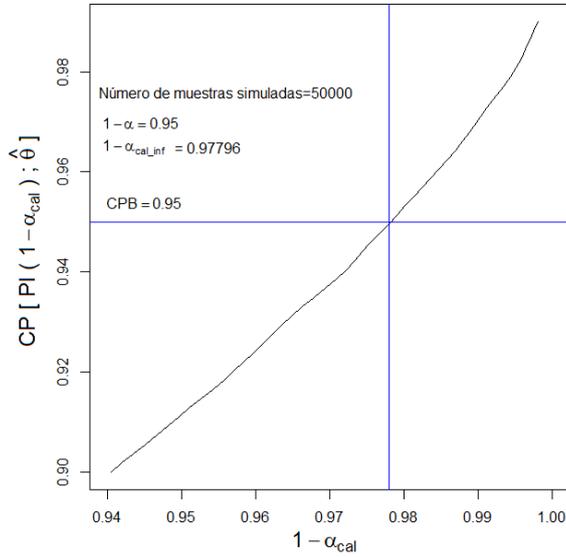


Figura 2.7: Curva de calibración del límite inferior por el Método 2 sin censura (CP=0.95)

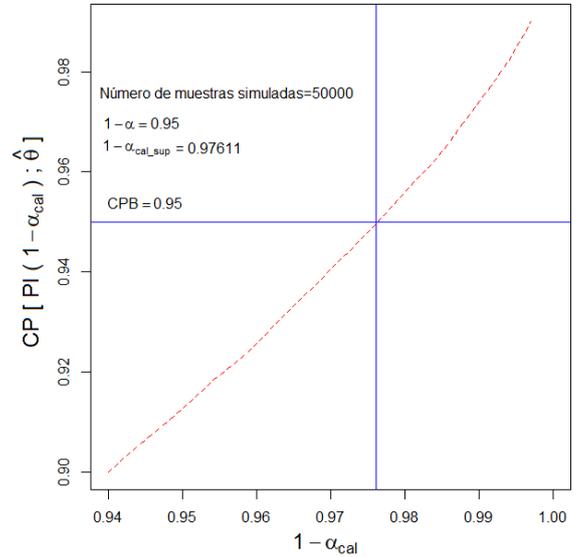


Figura 2.8: Curva de calibración del límite superior por el Método 2 sin censura (CP=0.95)

Usando los puntos de calibración con el Método 2 sin censura en la Figura 2.7 y en la Figura 2.8, se puede afirmar que un límite de predicción unilateral inferior naive al 97.80 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.61 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

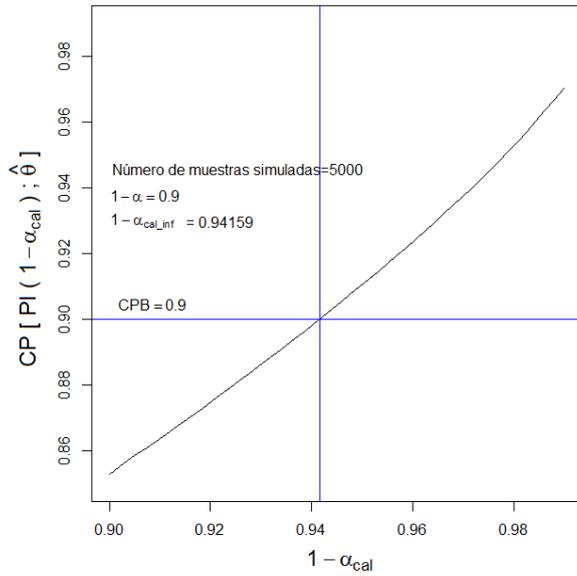


Figura 2.9: Curva de calibración del límite inferior por el Método 3 sin censura (CP=0.90)

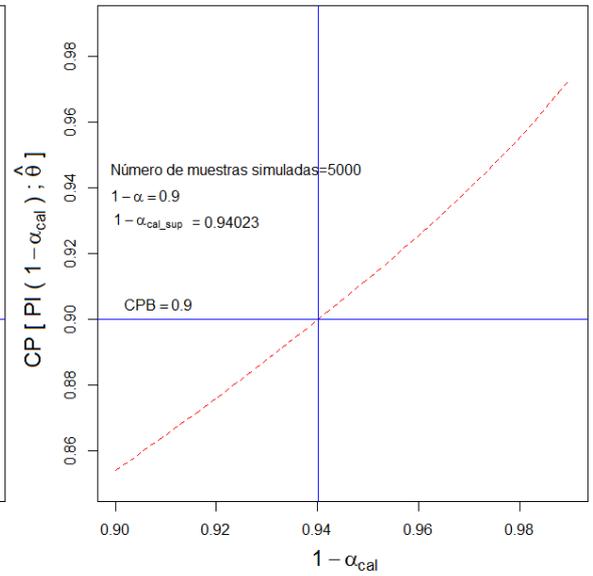


Figura 2.10: Curva de calibración del límite superior por el Método 3 sin censura (CP=0.90)

Usando los puntos de calibración con el Método 3 sin censura en la Figura 2.9 y en la Figura 2.10, se puede afirmar que un límite de predicción unilateral inferior naive al 94.16 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.02 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

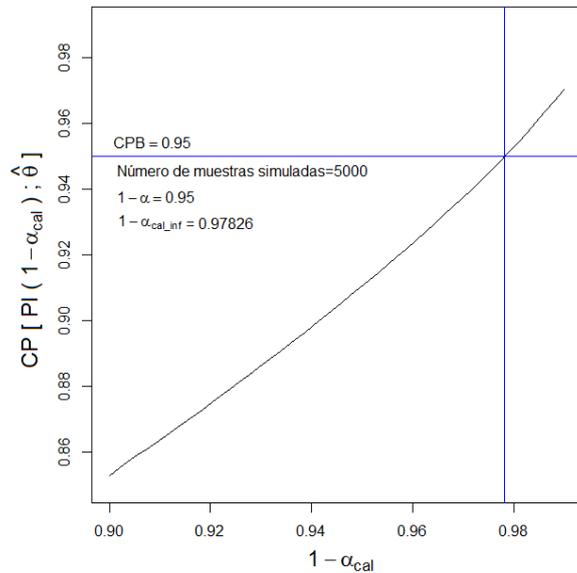


Figura 2.11: Curva de calibración del límite inferior por el Método 3 sin censura (CP=0.95)

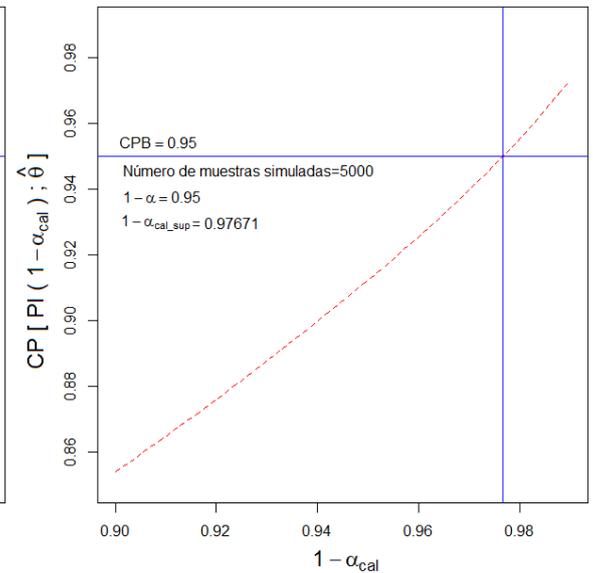


Figura 2.12: Curva de calibración del límite superior por el Método 3 sin censura (CP=0.95)

Usando los puntos de calibración con el Método 3 sin censura en la Figura 2.11 y en la Figura 2.12, se puede afirmar que un límite de predicción unilateral inferior naive al 97.83 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.67 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

A continuación se presentan los límites de predicción unilaterales calibrados para el tiempo de vida del componente A en campo sin censura:

Tabla 2.20: Límites de predicción unilaterales inferiores calibrados sin censura

CP[PI(1 - α_{cl})] = 0.90 Sin Censura			
Límite inferior naive	Límite inferior calibrado Método 1	Límite inferior calibrado Método 2	Límite inferior calibrado Método 3
[13.8146, +∞)	[10.885443, +∞)	[10.952989, +∞)	[10.870861, +∞)
CP[PI(1 - α_{cl})] = 0.95 Sin Censura			
[10.19789, +∞)	[7.496723, +∞)	[7.505484, +∞)	[7.470636, +∞)

Tabla 2.21: Límites de predicción unilaterales superiores calibrados sin censura

CP[PI(1 - α_{cu})] = 0.90 Sin Censura			
Límite superior naive	Límite superior calibrado Método 1	Límite superior calibrado Método 2	Límite superior calibrado Método 3
(0, 120.4639]	(0, 154.5540]	(0, 152.2468]	(0, 152.4882]
CP[PI(1 - α_{cu})] = 0.95 Sin Censura			
(0, 164.47993]	(0, 223.1378]	(0, 219.5695]	(0, 221.6355]

Para calcular un intervalo de predicción bilateral para el tiempo de vida del producto en campo al 90 % sin censura, los límites de predicción unilaterales naive al 95 % pueden ser combinados, como se muestra en la Tabla 2.22.

Tabla 2.22: Intervalos de predicción bilaterales calibrados sin censura

CP[PI(1 - α_c)] = 0.90 Sin Censura			
Intervalo naive	Intervalo calibrado Método 1	Intervalo calibrado Método 2	Intervalo calibrado Método 3
[10.19789, 164.47993]	[7.496723, 223.1378]	[7.505484, 219.5695]	[7.470636, 221.6355]

De acuerdo con estos resultados, con datos de vida sin censura, podemos estar 90 % seguros de que el componente A fallará entre la semana 7.50 y la semana 223.14 de uso calibrando con el Método 1, entre la semana 7.51 y la semana 219.57 calibrando con el Método 2 y entre la semana 7.5 y la semana 221.64 calibrando con el Método 3.

• **Curvas de calibración para datos de vida con Censura 1 (Percentil 95):**

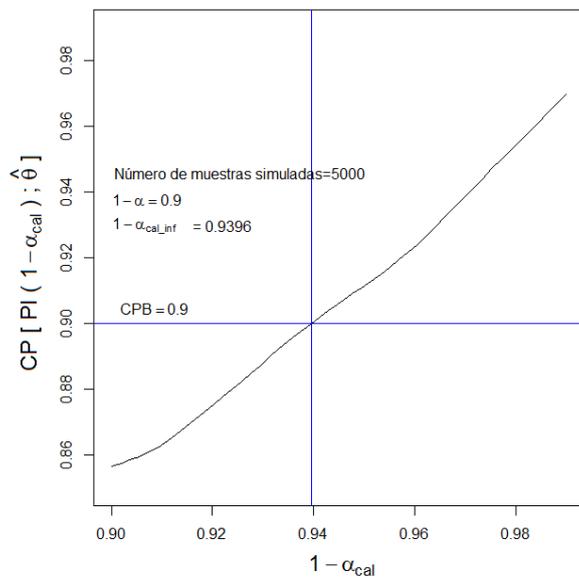


Figura 2.13: Curva de calibración del límite inferior por el Método 1 con Censura 1 (CP=0.90)

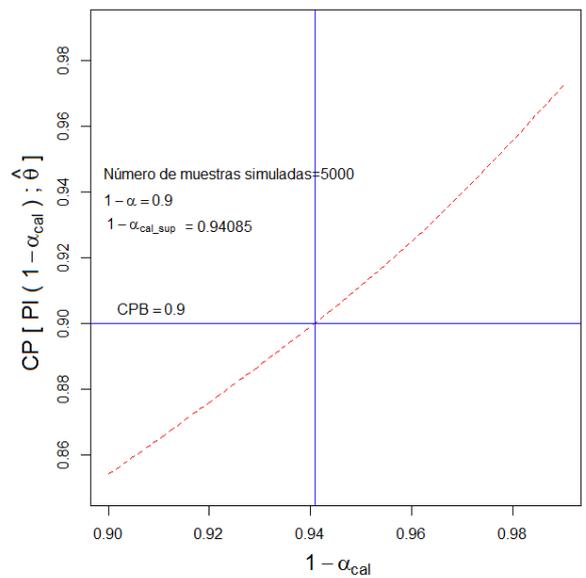


Figura 2.14: Curva de calibración del límite superior por el Método 1 con Censura 1 (CP=0.90)

Usando los puntos de calibración con el Método 1 con Censura 1 en la Figura 2.13 y en la Figura 2.14, se puede afirmar que un límite de predicción unilateral inferior naive al 93.96 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.09 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

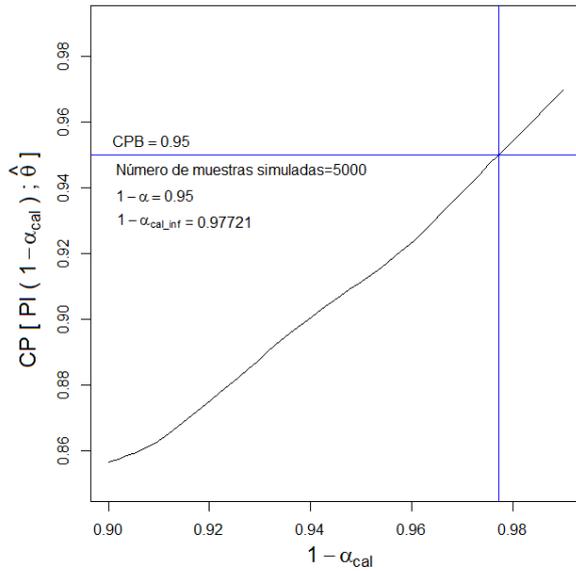


Figura 2.15: Curva de calibración del límite inferior por el Método 1 con Censura 1 (CP=0.95)

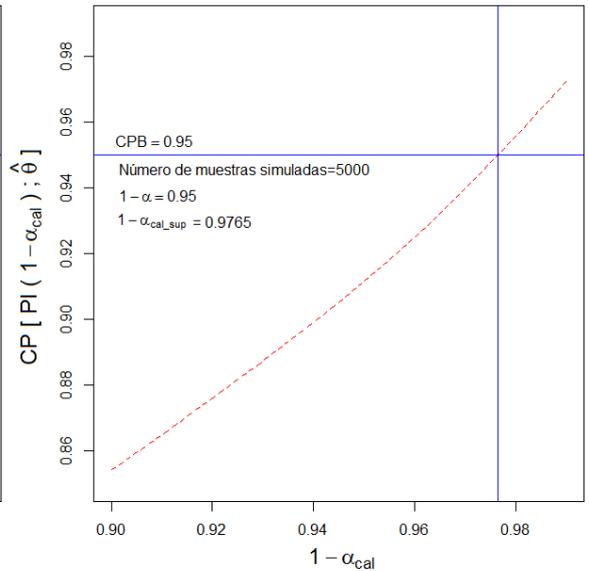


Figura 2.16: Curva de calibración del límite superior por el Método 1 con Censura 1 (CP=0.95)

Usando los puntos de calibración con el Método 1 con Censura 1 en la Figura 2.15 y en la Figura 2.16, se puede afirmar que un límite de predicción unilateral inferior naive al 97.72 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.65 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

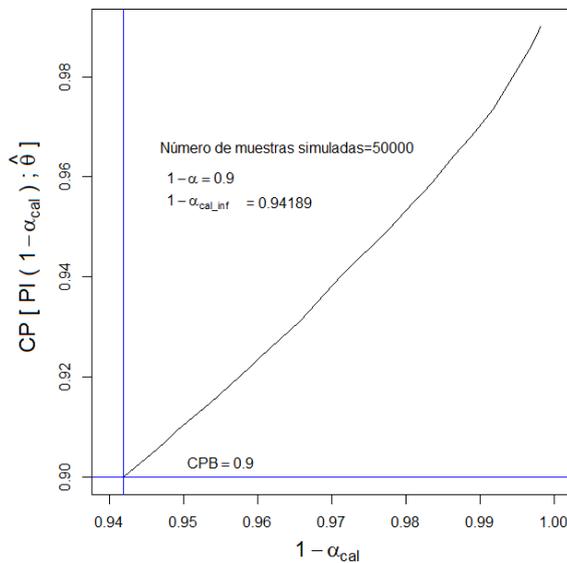


Figura 2.17: Curva de calibración del límite inferior por el Método 2 con Censura 1 (CP=0.90)

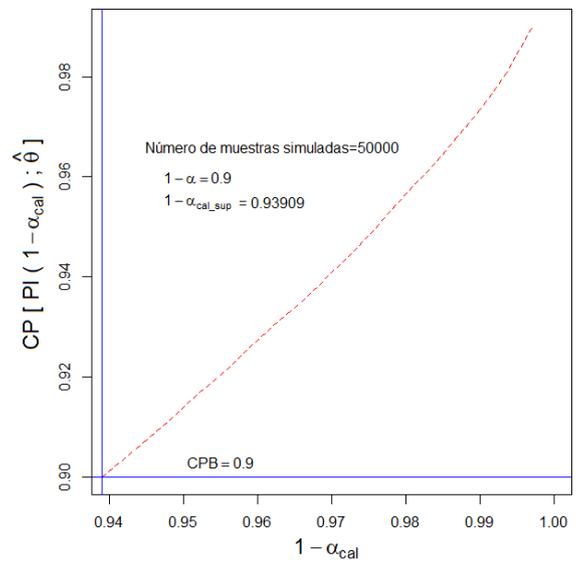


Figura 2.18: Curva de calibración del límite superior por el Método 2 con Censura 1 (CP=0.90)

Usando los puntos de calibración con el Método 2 con Censura 1 en la Figura 2.17 y en la Figura 2.18, se puede afirmar que un límite de predicción unilateral inferior naive al 94.19 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 93.91 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

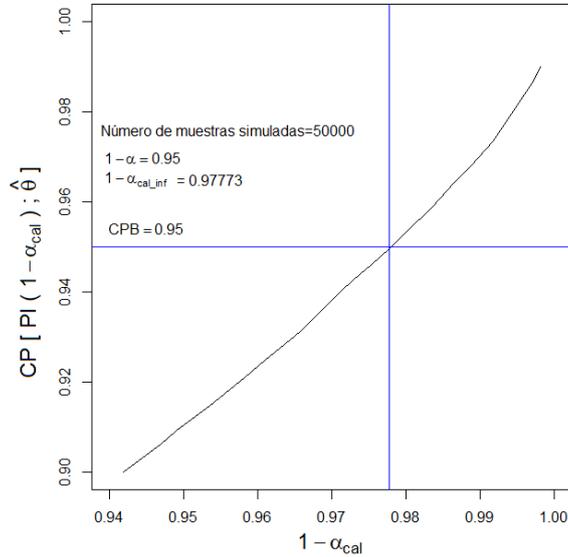


Figura 2.19: Curva de calibración del límite inferior por el Método 2 con Censura 1 (CP=0.95)

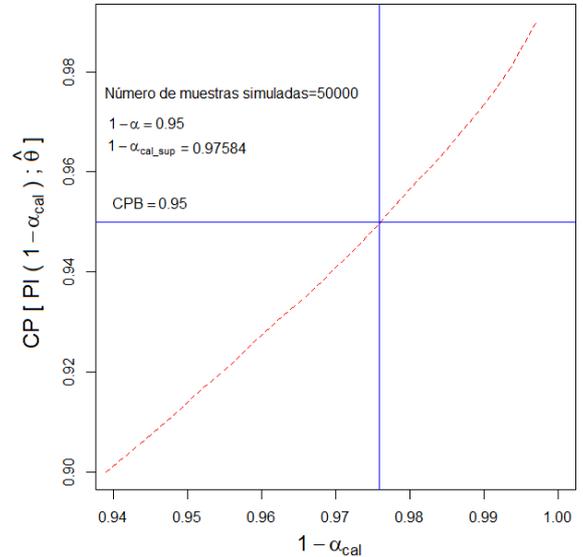


Figura 2.20: Curva de calibración del límite superior por el Método 2 con Censura 1 (CP=0.95)

Usando los puntos de calibración con el Método 2 con Censura 1 en la Figura 2.19 y en la Figura 2.20, se puede afirmar que un límite de predicción unilateral inferior naive al 97.77 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.58 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

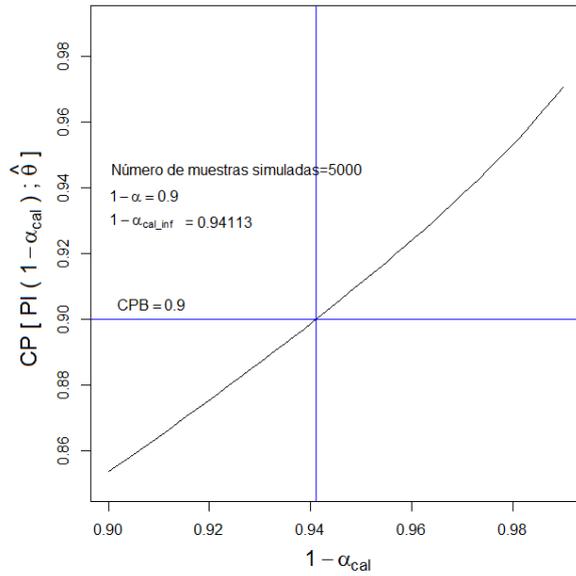


Figura 2.21: Curva de calibración del límite inferior por el Método 3 con Censura 1 (CP=0.90)

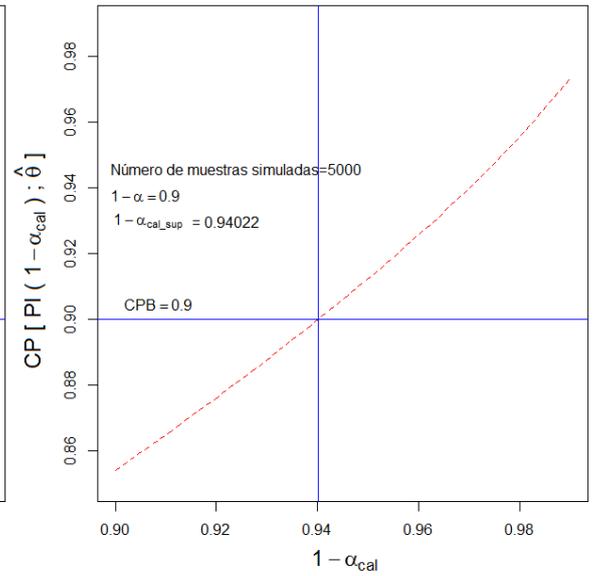


Figura 2.22: Curva de calibración del límite superior por el Método 3 con Censura 1 (CP=0.90)

Usando los puntos de calibración con el Método 3 con Censura 1 en la Figura 2.21 y en la Figura 2.22, se puede afirmar que un límite de predicción unilateral inferior naive al 94.11 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.02 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

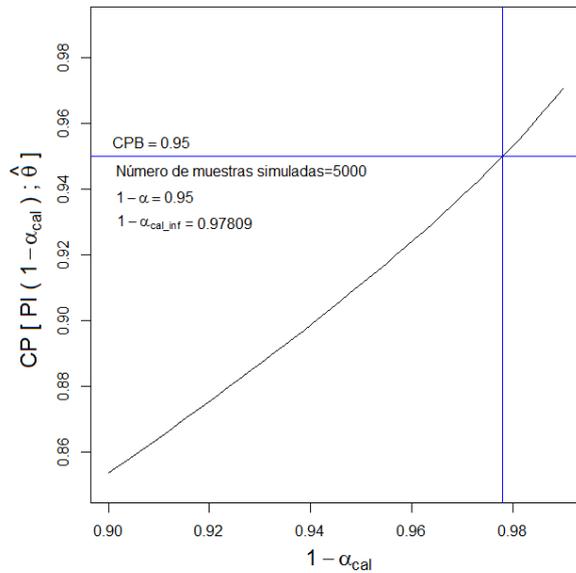


Figura 2.23: Curva de calibración del límite inferior por el Método 3 con Censura 1 (CP=0.95)

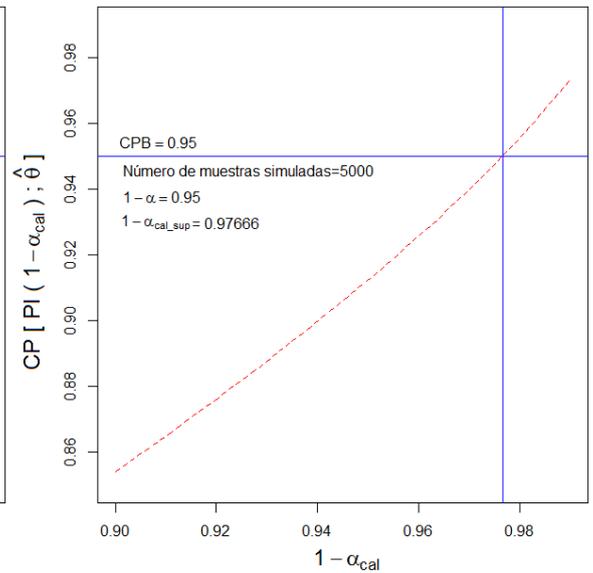


Figura 2.24: Curva de calibración del límite superior por el Método 3 con Censura 1 (CP=0.95)

Usando los puntos de calibración con el Método 3 con Censura 1 en la Figura 2.23 y en la Figura 2.24, se puede afirmar que un límite de predicción unilateral inferior naive al 97.81 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.67 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

A continuación se presentan los límites de predicción unilaterales calibrados para el tiempo de vida del componente A en campo con Censura 1:

Tabla 2.23: Límites de predicción unilaterales inferiores calibrados con Censura 1 (Percentil 95)

CP[PI(1 - α_{cl})] = 0.90 Con Censura 1			
Límite inferior naive	Límite inferior calibrado Método 1	Límite inferior calibrado Método 2	Límite inferior calibrado Método 3
[13.81653, +∞)	[11.046063, +∞)	[10.870159, +∞)	[10.928992, +∞)
CP[PI(1 - α_{cl})] = 0.95 Con Censura 1			
[10.22435, +∞)	[7.632642, +∞)	[7.571853, +∞)	[7.528413, +∞)

Tabla 2.24: Límites de predicción unilaterales superiores calibrados con Censura 1 (Percentil 95)

CP[PI(1 - α_{cu})] = 0.90 Con Censura 1			
Límite superior naive	Límite superior calibrado Método 1	Límite superior calibrado Método 2	Límite superior calibrado Método 3
(0, 118.45863]	(0, 150.3480]	(0, 148.4723]	(0, 149.6644]
CP[PI(1 - α_{cu})] = 0.95 Con Censura 1			
(0, 161.36658]	(0, 216.2695]	(0, 214.0658]	(0, 216.7886]

Para calcular un intervalo de predicción bilateral para el tiempo de vida del producto en campo al 90 % con Censura 1, los límites de predicción unilaterales naive al 95 % pueden ser combinados, como se muestra en la Tabla 2.25.

Tabla 2.25: Intervalos de predicción bilaterales calibrados con Censura 1

CP[PI(1 - α_c)] = 0.90 Con Censura 1			
Intervalo naive	Intervalo calibrado Método 1	Intervalo calibrado Método 2	Intervalo calibrado Método 3
[10.22435, 161.36658]	[7.632642, 216.2695]	[7.571853, 214.0658]	[7.528413, 216.7886]

De acuerdo con estos resultados, con datos de vida con Censura 1, podemos estar 90% seguros de que el componente A fallará entre la semana 7.63 y la semana 216.27 de uso calibrando con el Método 1, entre la semana 7.57 y la semana 214.07 calibrando con el Método 2 y entre la semana 7.53 y la semana 216.79 calibrando con el Método 3.

• **Curvas de calibración para datos de vida con Censura 2 (Percentil 80):**

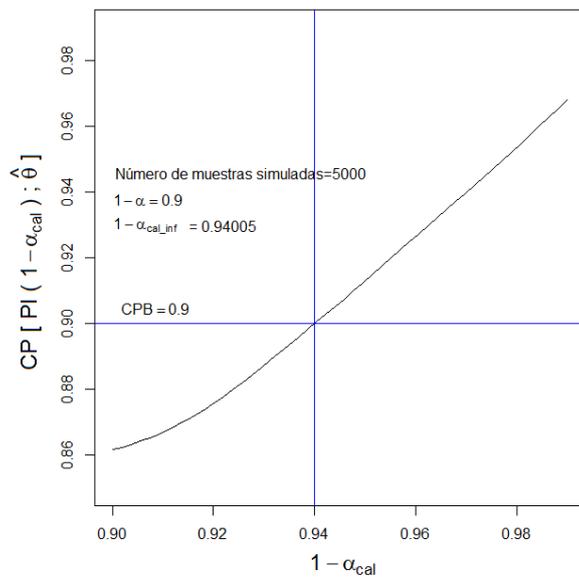


Figura 2.25: Curva de calibración del límite inferior por el Método 1 con Censura 2 (CP=0.90)

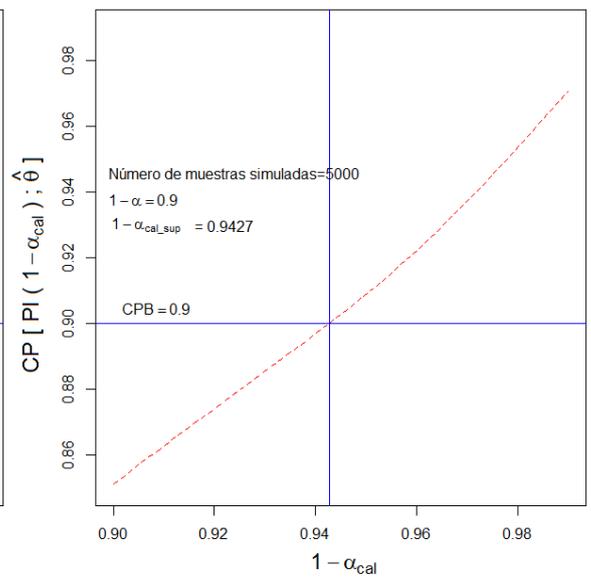


Figura 2.26: Curva de calibración del límite superior por el Método 1 con Censura 2 (CP=0.90)

Usando los puntos de calibración con el Método 1 con Censura 2 en la Figura 2.25 y en la Figura 2.26, se puede afirmar que un límite de predicción unilateral inferior naive al 94.01% proporciona un límite de predicción unilateral inferior calibrado aproximado al 90% para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.27% proporciona un límite de predicción unilateral superior calibrado aproximado al 90% para el tiempo de vida del producto en campo.

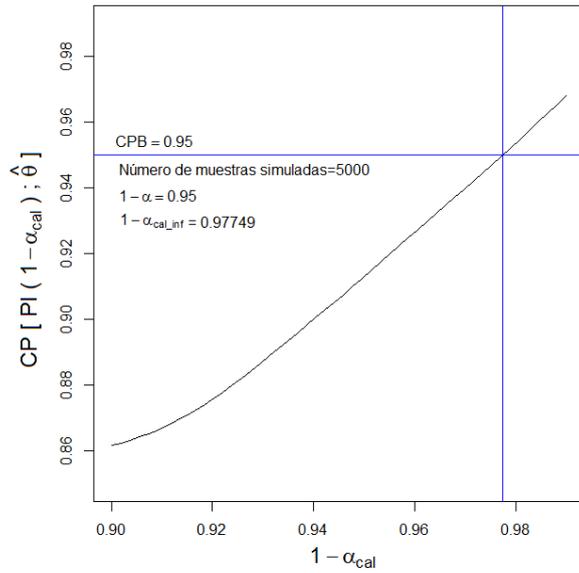


Figura 2.27: Curva de calibración del límite inferior por el Método 1 con Censura 2 (CP=0.95)

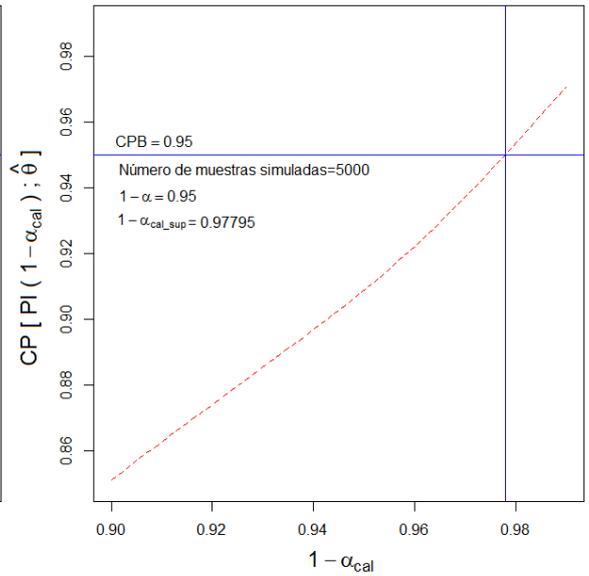


Figura 2.28: Curva de calibración del límite superior por el Método 1 con Censura 2 (CP=0.95)

Usando los puntos de calibración con el Método 1 con Censura 2 en la Figura 2.27 y en la Figura 2.28, se puede afirmar que un límite de predicción unilateral inferior naive al 97.75 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.80 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

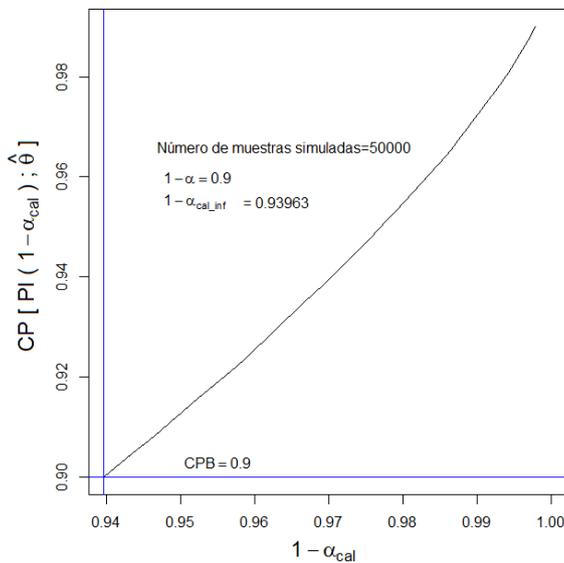


Figura 2.29: Curva de calibración del límite inferior por el Método 2 con Censura 2 (CP=0.90)

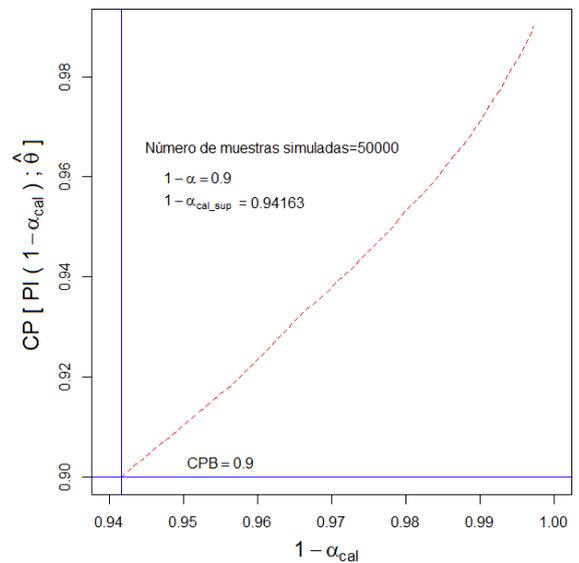


Figura 2.30: Curva de calibración del límite superior por el Método 2 con Censura 2 (CP=0.90)

Usando los puntos de calibración con el Método 2 con Censura 2 en la Figura 2.29 y en la Figura 2.30, se puede afirmar que un límite de predicción unilateral inferior naive al 93.96 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.16 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

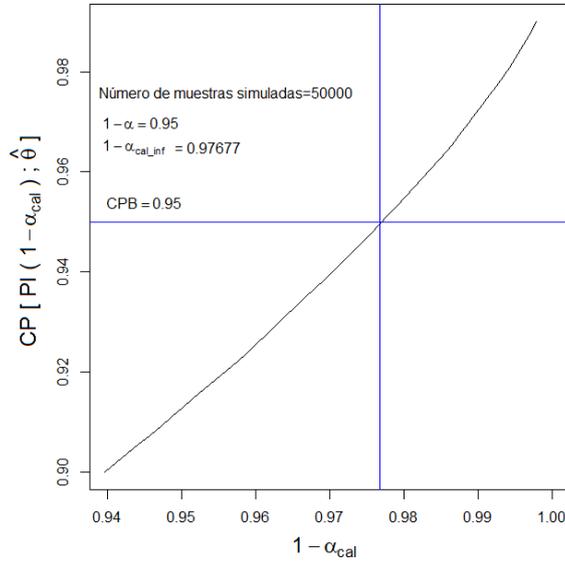


Figura 2.31: Curva de calibración del límite inferior por el Método 2 con Censura 2 (CP=0.95)

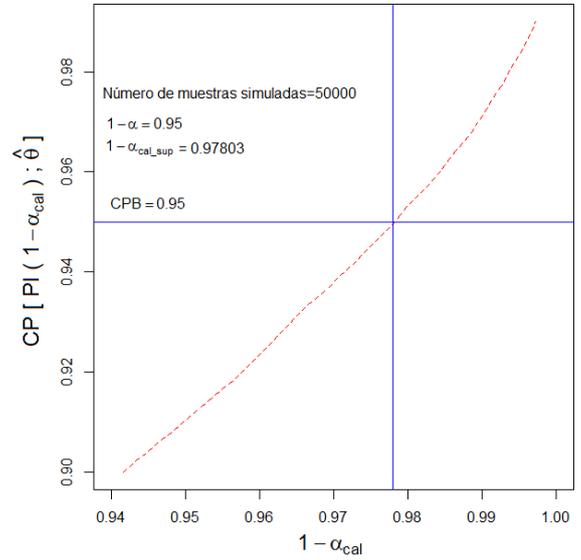


Figura 2.32: Curva de calibración del límite superior por el Método 2 con Censura 2 (CP=0.95)

Usando los puntos de calibración con el Método 2 con Censura 2 en la Figura 2.31 y en la Figura 2.32, se puede afirmar que un límite de predicción unilateral inferior naive al 97.68 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.80 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

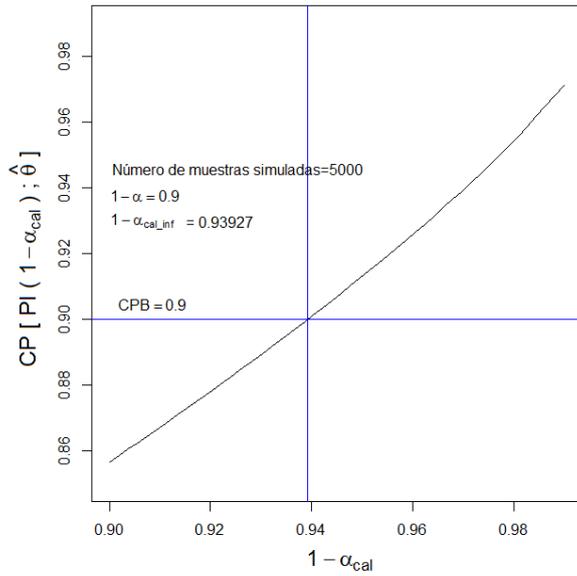


Figura 2.33: Curva de calibración del límite inferior por el Método 3 con Censura 2 (CP=0.90)

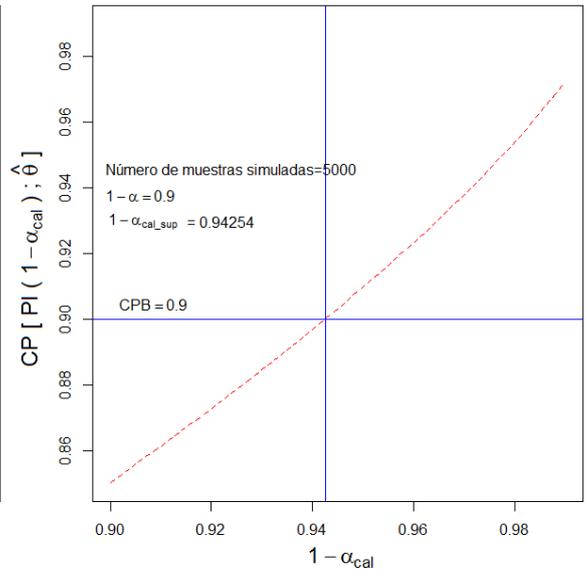


Figura 2.34: Curva de calibración del límite superior por el Método 3 con Censura 2 (CP=0.90)

Usando los puntos de calibración con el Método 3 con Censura 2 en la Figura 2.33 y en la Figura 2.34, se puede afirmar que un límite de predicción unilateral inferior naive al 93.93 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.25 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

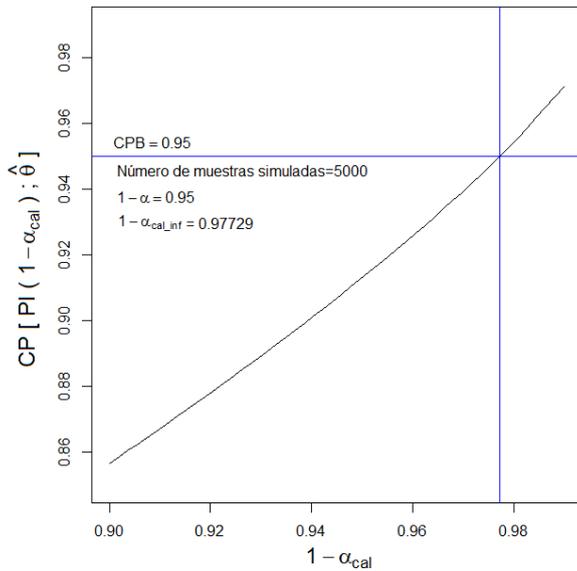


Figura 2.35: Curva de calibración del límite inferior por el Método 3 con Censura 2 (CP=0.95)

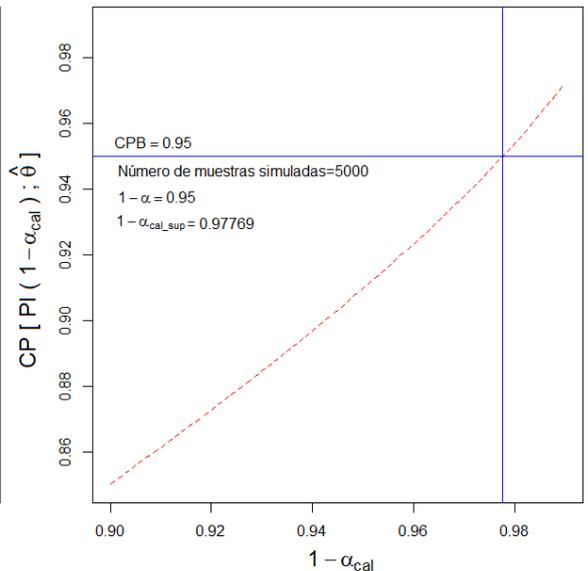


Figura 2.36: Curva de calibración del límite superior por el Método 3 con Censura 2 (CP=0.95)

Usando los puntos de calibración con el Método 3 con Censura 2 en la Figura 2.35 y en la Figura 2.36, se puede afirmar que un límite de predicción unilateral inferior naive al 97.73 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.77 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

A continuación se presentan los límites de predicción unilaterales calibrados para el tiempo de vida del componente A en campo con Censura 2:

Tabla 2.26: Límites de predicción unilaterales inferiores calibrados con Censura 2 (Percentil 80)

CP[PI(1 - α_{cl})] = 0.90 Con Censura 2			
Límite inferior naive	Límite inferior calibrado Método 1	Límite inferior calibrado Método 2	Límite inferior calibrado Método 3
[13.53302, + ∞)	[10.816064, + ∞)	[10.847256, + ∞)	[10.874441, + ∞)
CP[PI(1 - α_{cl})] = 0.95 Con Censura 2			
[10.05261, + ∞)	[7.500651, + ∞)	[7.581461, + ∞)	[7.522553, + ∞)

Tabla 2.27: Límites de predicción unilaterales superiores con Censura 2 (Percentil 80)

CP[PI(1 - α_{cu})] = 0.90 Con Censura 2			
Límite superior naive	Límite superior calibrado Método 1	Límite superior calibrado Método 2	Límite superior calibrado Método 3
(0, 113.02918]	(0, 144.9831]	(0, 143.8535]	(0, 144.8131]
CP[PI(1 - α_{cu})] = 0.95 Con Censura 2			
(0, 153.41521]	(0, 209.6565]	(0, 209.9436]	(0, 208.7856]

Para calcular un intervalo de predicción bilateral para el tiempo de vida del producto en campo al 90 % con Censura 2, los límites de predicción unilaterales naive al 95 % pueden ser combinados, como se muestra en la Tabla 2.28.

Tabla 2.28: Intervalos de predicción bilaterales calibrados con Censura 2

CP[PI(1 - α_c)] = 0.90 Con Censura 2			
Intervalo naive	Intervalo calibrado Método 1	Intervalo calibrado Método 2	Intervalo calibrado Método 3
[10.05261, 153.41521]	[7.500651, 209.6565]	[7.581461, 209.9436]	[7.522553, 208.7856]

De acuerdo con estos resultados, con datos de vida con Censura 2, podemos estar 90% seguros de que el componente A fallará entre la semana 7.50 y la semana 209.66 de uso calibrando con el Método 1, entre la semana 7.58 y la semana 209.94 calibrando con el Método 2 y entre la semana 7.52 y la semana 208.79 calibrando con el Método 3.

• **Curvas de calibración para datos de vida con Censura 3 (Percentil 70):**

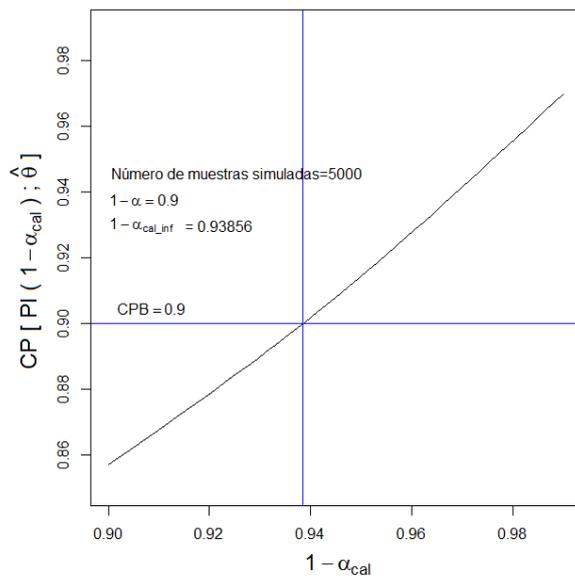


Figura 2.37: Curva de calibración del límite inferior por el Método 1 con Censura 3 (CP=0.90)

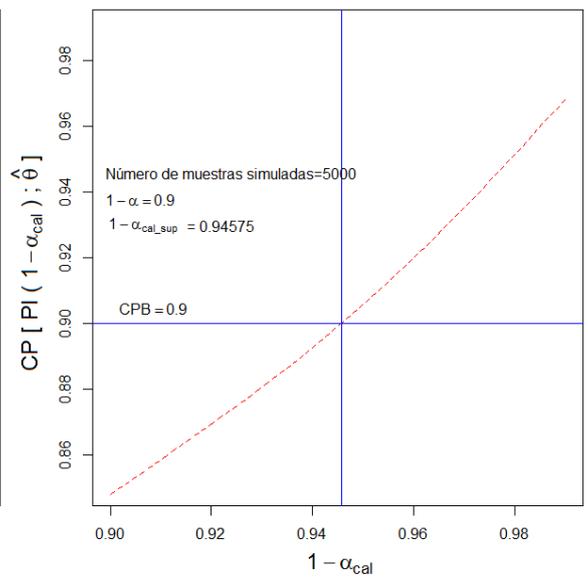


Figura 2.38: Curva de calibración del límite superior por el Método 1 con Censura 3 (CP=0.90)

Usando los puntos de calibración con el Método 1 con Censura 3 en la Figura 2.37 y en la Figura 2.38, se puede afirmar que un límite de predicción unilateral inferior naive al 93.86% proporciona un límite de predicción unilateral inferior calibrado aproximado al 90% para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.58% proporciona un límite de predicción unilateral superior calibrado aproximado al 90% para el tiempo de vida del producto en campo.

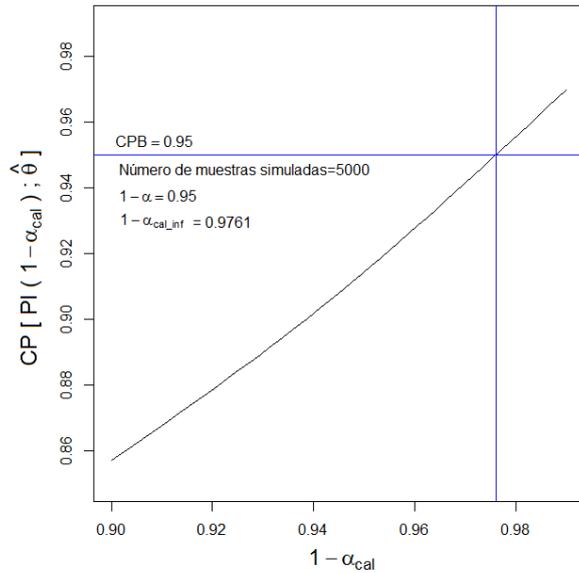


Figura 2.39: Curva de calibración del límite inferior por el Método 1 con Censura 3 (CP=0.95)

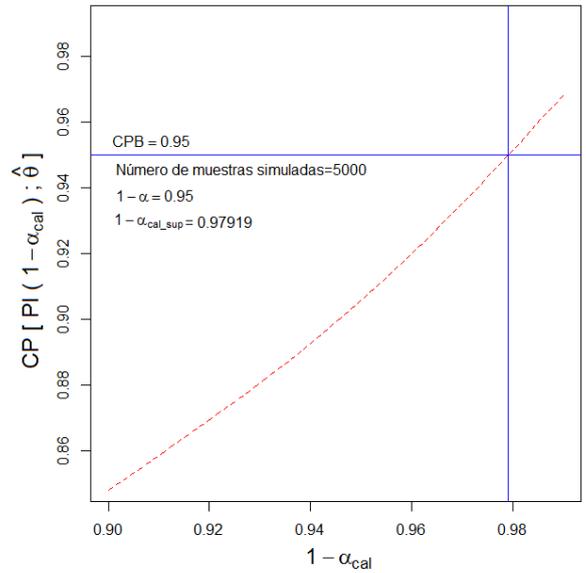


Figura 2.40: Curva de calibración del límite superior por el Método 1 con Censura 3 (CP=0.95)

Usando los puntos de calibración con el Método 1 con Censura 3 en la Figura 2.39 y en la Figura 2.40, se puede afirmar que un límite de predicción unilateral inferior naive al 97.61 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.92 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

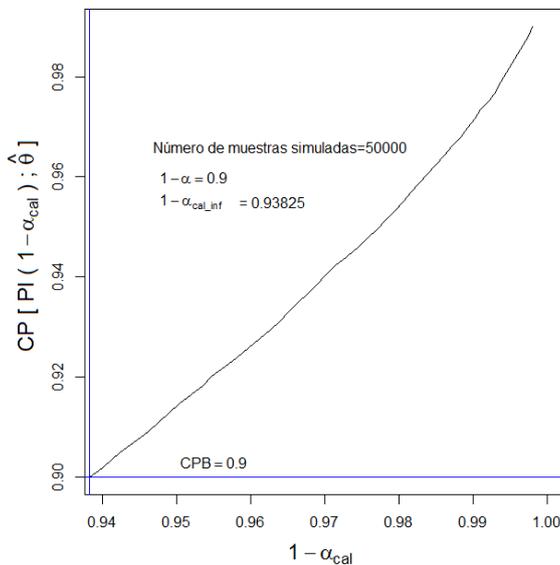


Figura 2.41: Curva de calibración del límite inferior por el Método 2 con Censura 3 (CP=0.90)

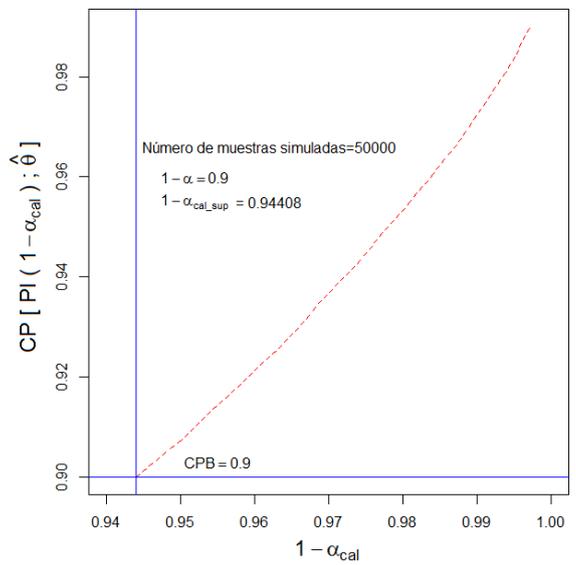


Figura 2.42: Curva de calibración del límite superior por el Método 2 con Censura 3 (CP=0.90)

Usando los puntos de calibración con el Método 2 con Censura 3 en la Figura 2.41 y en la Figura 2.42, se puede afirmar que un límite de predicción unilateral inferior naive al 93.83 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.41 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

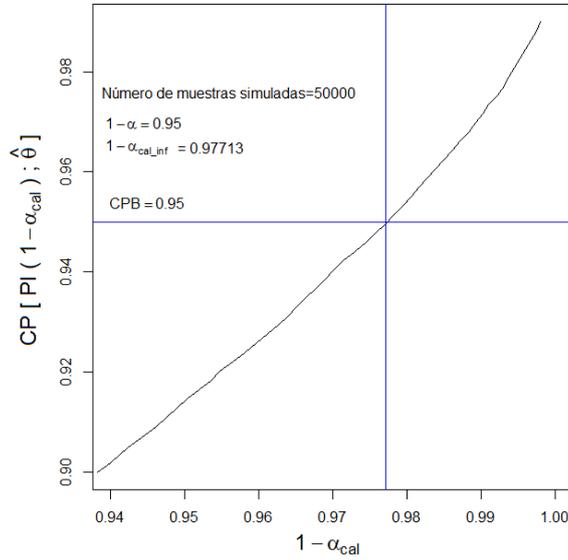


Figura 2.43: Curva de calibración del límite inferior por el Método 2 con Censura 3 (CP=0.95)

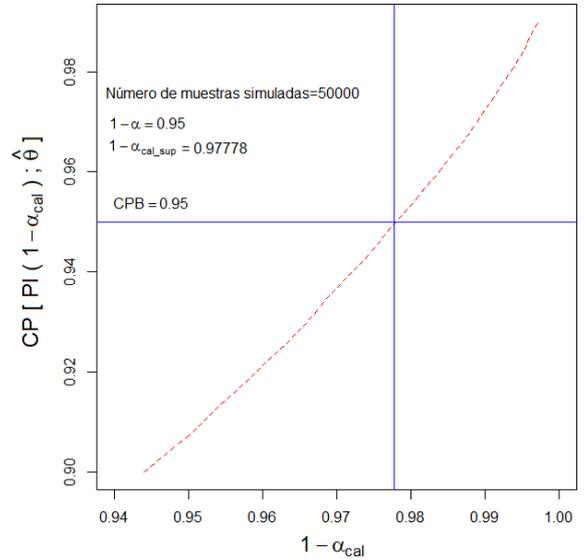


Figura 2.44: Curva de calibración del límite superior por el Método 2 con Censura 3 (CP=0.95)

Usando los puntos de calibración con el Método 2 con Censura 3 en la Figura 2.43 y en la Figura 2.44, se puede afirmar que un límite de predicción unilateral inferior naive al 97.71 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.78 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

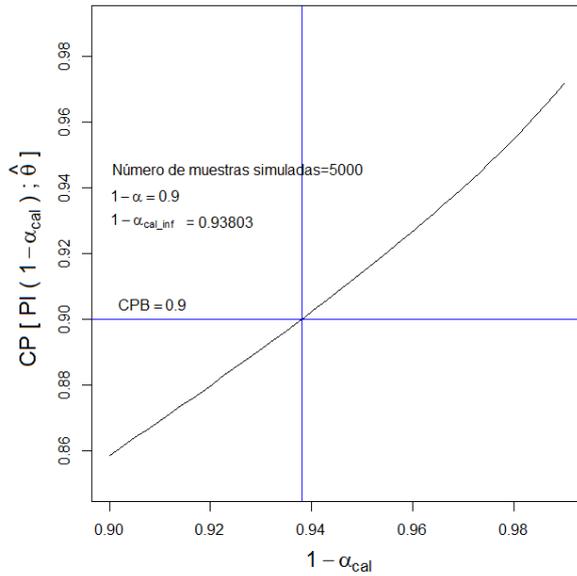


Figura 2.45: Curva de calibración del límite inferior por el Método 3 con Censura 3 (CP=0.90)

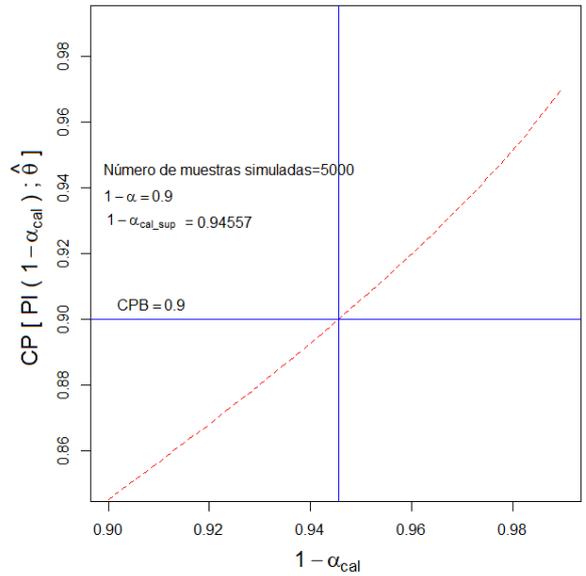


Figura 2.46: Curva de calibración del límite superior por el Método 3 con Censura 3 (CP=0.90)

Usando los puntos de calibración con el Método 3 con Censura 3 en la Figura 2.45 y en la Figura 2.46, se puede afirmar que un límite de predicción unilateral inferior naive al 93.80 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 94.56 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

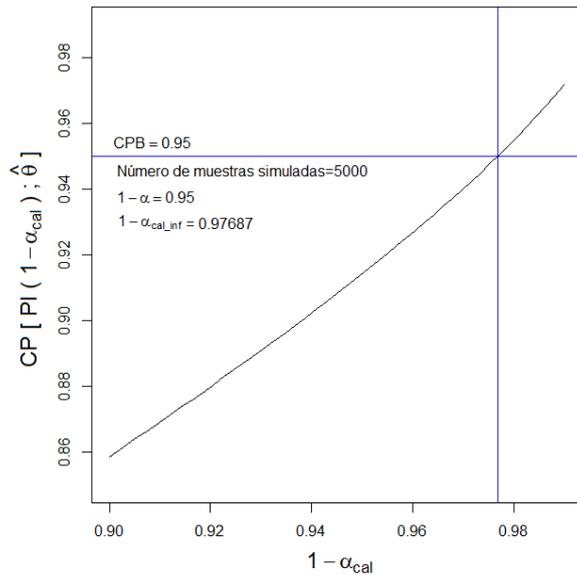


Figura 2.47: Curva de calibración del límite inferior por el Método 3 con Censura 3 (CP=0.95)

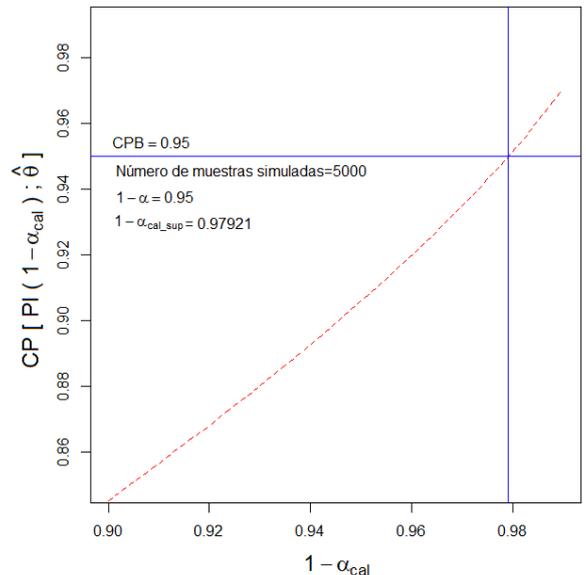


Figura 2.48: Curva de calibración del límite superior por el Método 3 con Censura 3 (CP=0.95)

Usando los puntos de calibración con el Método 3 con Censura 3 en la Figura 2.47 y en la Figura 2.48, se puede afirmar que un límite de predicción unilateral inferior naive al 97.67 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 97.92 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

A continuación se presentan los límites de predicción unilaterales calibrados para el tiempo de vida del componente A en campo con Censura 3:

Tabla 2.29: Límites de predicción unilaterales inferiores calibrados con Censura 3 (Percentil 70)

CP[PI(1 - α_{cl})] = 0.90 Con Censura 3			
Límite inferior naive	Límite inferior calibrado Método 1	Límite inferior calibrado Método 2	Límite inferior calibrado Método 3
[13.26696, + ∞)	[10.691176, + ∞)	[10.713769, + ∞)	[10.729732, + ∞)
CP[PI(1 - α_{cl})] = 0.95 Con Censura 3			
[9.828942, + ∞)	[7.466977, + ∞)	[7.354432, + ∞)	[7.382321, + ∞)

Tabla 2.30: Límites de predicción unilaterales superiores calibrados con Censura 3 (Percentil 70)

CP[PI(1 - α_{cu})] = 0.90 Con Censura 3			
Límite superior naive	Límite superior calibrado Método 1	Límite superior calibrado Método 2	Límite superior calibrado Método 3
(0, 112.84895]	(0, 148.4274]	(0, 146.5511]	(0, 148.2265]
CP[PI(1 - α_{cu})] = 0.95 Con Censura 3			
(0, 153.556945]	(0, 214.7998]	(0, 209.8132]	(0, 214.8834]

Para calcular un intervalo de predicción bilateral para el tiempo de vida del producto en campo al 90 % con Censura 3, los límites de predicción unilaterales naive al 95 % pueden ser combinados, como se muestra en la Tabla 2.31.

Tabla 2.31: Intervalos de predicción bilaterales calibrados con Censura 3

CP[PI(1 - α_c)] = 0.90 Con Censura 3			
	Intervalo calibrado Método 1	Intervalo calibrado Método 2	Intervalo calibrado Método 3
Intervalo naive	[9.828942, 153.556945]	[7.466977, 214.7998]	[7.581461, 209.8132]
	[7.522553, 214.8834]		

De acuerdo con estos resultados, con datos de vida con Censura 3, podemos estar 90% seguros de que el componente A fallará entre la semana 7.47 y la semana 214.80 de uso calibrando con el Método 1, entre la semana 7.58 y la semana 209.81 calibrando con el Método 2 y entre la semana 7.52 y la semana 214.88 calibrando con el Método 3.

Por otro lado, es posible determinar las características de los intervalos de predicción en términos de su longitud promedio, calculada como la diferencia entre el límite superior y el límite inferior de cada intervalo. En la Tabla 2.32 se presentan las longitudes promedio para los intervalos bilaterales naive y calibrados con los diferentes tiempos de censura y para $CP[PI(1 - \alpha_c)] = 0.90$.

Tabla 2.32: Longitud de los Intervalos de predicción bilaterales calibrados

CP[PI(1 - α_c)] = 0.90 Sin Censura			
	Longitud intervalo calibrado Método 1	Longitud intervalo calibrado Método 2	Longitud intervalo calibrado Método 3
Longitud intervalo naive	154.2820	215.6411	212.0640
			214.1649
CP[PI(1 - α_c)] = 0.90 Con Censura 1			
	151.1422	208.6369	206.4939
			209.2602
CP[PI(1 - α_c)] = 0.90 Con Censura 2			
	143.3626	202.1558	202.3621
			201.2630
CP[PI(1 - α_c)] = 0.90 Con Censura 3			
	143.7280	207.3328	202.2371
			207.3608

En general, de estos resultados se puede inferir que a medida que disminuyen los tiempos de censura, los intervalos tienden a ser más estrechos, esto puede explicarse por el hecho de que la estimación puntual dada por $\hat{\eta}$ también toma un menor valor por lo que se puede pensar que se produce una subestimación, lo cual ocurre análogamente con los límites de

los intervalos, más específicamente con el límite superior. Con respecto a las metodologías de calibración para los límites de predicción naive propuestas por Escobar y Meeker (1999), se puede decir que los resultados arrojados por el Método 1 y Método 2, que son casos particulares de la calibración por simulación del proceso de muestreo/predicción y por el Método 3, donde la calibración se obtiene promediando las probabilidades de cobertura condicional, son muy similares, sin embargo, de acuerdo con la longitud de los intervalos calibrados obtenidos y teniendo en cuenta que tienen la misma probabilidad de cobertura que es 90 %, el método que proporciona los mejores resultados es el Método 2, ya que los intervalos son más cortos, además coincide con el hecho de que el algoritmo por medio del cual se lleva a cabo este procedimiento es computacionalmente más eficiente que el de las otras dos metodologías, esto debido a que los tiempos de las simulaciones son menores aún cuando los tamaños muestrales fueron mayores ($B = 50000$).

Predicción estadística para el componente A con información sólo de la ALT

Resulta interesante evaluar la pertinencia del modelo de tasa de uso propuesto Meeker *et al.* (2009) en el momento de realizar predicción estadística para el tiempo de vida del componente A en campo. Debido a que este modelo integra los datos obtenidos a partir de las pruebas de laboratorio con la información complementaria de las reclamaciones por garantía y de las condiciones de uso del producto, para definir la utilidad de dicho modelo, se omitirá la información de la distribución de las tasas de uso del producto en manos del cliente estimada por el departamento de marketing de la compañía y se tomará un único valor como tasa de uso que será la que se presenta con mayor frecuencia, que es de 4 usos por semana (ver Figura 1.1). También se omitirá la información proveniente de las devoluciones por garantía, por lo que la distribución del tiempo de vida del componente A estará dada por

$$F_C(c) = \Pr(C \leq c) = \Phi_{nor} \left[\frac{\log((c/R)/\eta_C)}{\sigma_C} \right] = \Phi_{nor} \left[\frac{\log(c) - \log(\eta_C \times R)}{\sigma_C} \right]$$

Teniendo en cuenta lo anterior, se realizará todo el proceso de predicción sólo para el caso con datos de vida sin censura. La programación puede verse en el **apéndice A**.

Los parámetros estimados para los datos de vida sin censura en unidades de los ciclos hasta la falla en la prueba de laboratorio se muestran en la Tabla 3.1.

Tabla 3.1: Parámetros estimados vía Evolución Diferencial para los datos de vida del componente A sin censura sólo con la ALT.

Parámetro	Estimación
$\hat{\eta}_C$	186.4633196
$\hat{\sigma}_C$	0.7447178

A continuación se presentan los intervalos de predicción bilaterales y los límites de predicción unilaterales naive para el tiempo de vida del componente A en campo usando sólo la distribución de tiempo de vida dada por la ALT con datos sin censura y asumiendo que la tasa de uso del producto es constante dentro de la población de consumidores, es decir, de 4 usos por semana. Además se presentan los resultados obtenidos utilizando el modelo de tasa de uso para poder hacer las respectivas comparaciones. Es necesario tener en cuenta que las unidades para éstos intervalos y límites están en términos de las semanas de servicio del producto.

Tabla 3.2: Intervalos de predicción bilaterales naive sin censura sólo con la ALT

Sin Censura		
CP[PI(1 - α); $\hat{\theta}$]	Modelo de tasa de uso PI(1 - α)	ALT PI(1 - α)
0.90	[10.19789, 164.47993]	[219.1107, 2538.8860]
0.95	[7.845625, 215.916508]	[173.280, 3210.394]

Tabla 3.3: Límites de predicción unilaterales inferiores naive sin censura sólo con la ALT

Sin Censura		
CP[PI(1 - α_l); $\hat{\theta}$]	Modelo de tasa de uso PI(1 - α_l)	ALT PI(1 - α_l)
0.90	[13.8146, $+\infty$)	[287.1873, $+\infty$)
0.95	[10.19789, $+\infty$)	[219.1107, $+\infty$)

Tabla 3.4: Límites de predicción unilaterales superiores naive sin censura sólo con la ALT

Sin Censura		
$CP[PI(1 - \alpha_u); \hat{\theta}]$	Modelo de tasa de uso $PI(1 - \alpha_u)$	ALT $PI(1 - \alpha_u)$
0.90	(0, 120.4639]	(0, 1937.0533]
0.95	(0, 164.47993]	(0, 2538.8860]

Para la calibración de los límites de predicción naive se utilizará sólo el Método 1, debido a que los resultados dados por los tres métodos son muy similares y el objetivo radica en identificar las diferencias entre los modelos para la predicción.

- **Curvas de calibración para datos de vida sin censura provenientes de la ALT:**

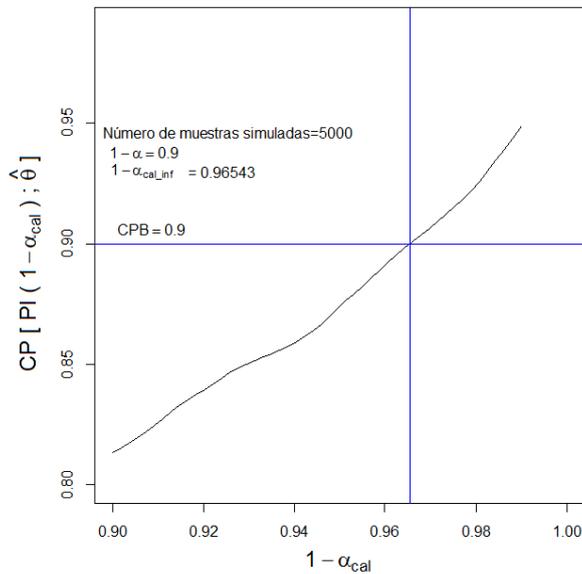


Figura 3.1: Curva de calibración del límite inferior por el Método 1 sin censura en la ALT (CP=0.90)

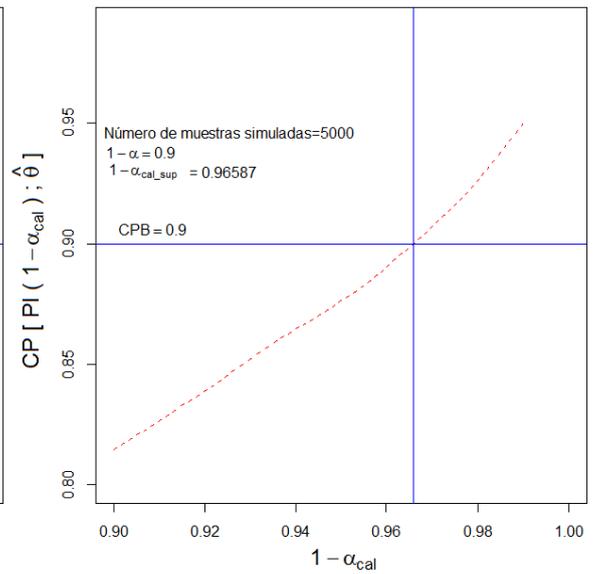


Figura 3.2: Curva de calibración del límite superior por el Método 1 sin censura en la ALT (CP=0.90)

Usando los puntos de calibración con el Método 1 sin censura en la Figura 3.1 y en la Figura 3.2, se puede afirmar que un límite de predicción unilateral inferior naive al 96.54 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 90 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral

superior naive al 96.59 % proporciona un límite de predicción unilateral superior calibrado aproximado al 90 % para el tiempo de vida del producto en campo.

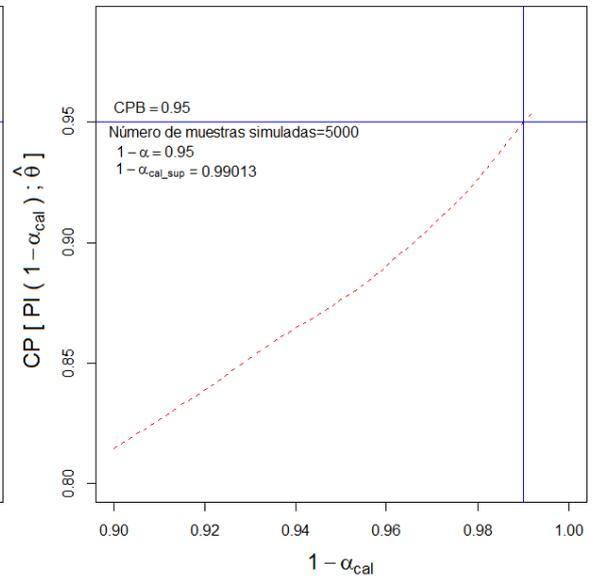
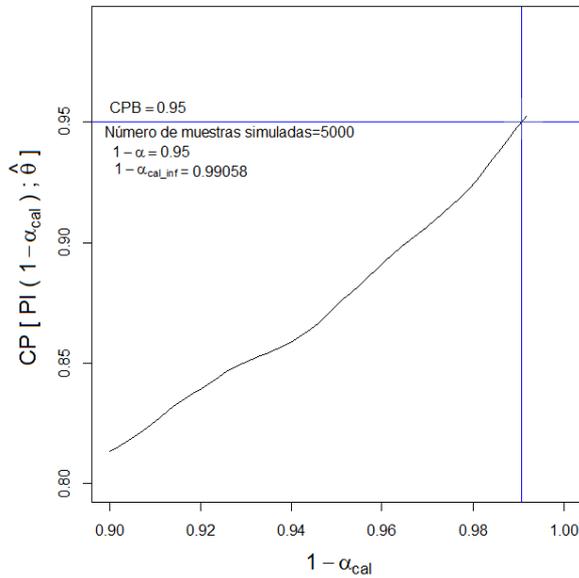


Figura 3.3: Curva de calibración del límite inferior por el Método 1 sin censura en la ALT (CP=0.95)

Figura 3.4: Curva de calibración del límite superior por el Método 1 sin censura en la ALT (CP=0.95)

Usando los puntos de calibración con el Método 1 sin censura en la Figura 3.3 y en la Figura 3.4, se puede afirmar que un límite de predicción unilateral inferior naive al 99.06 % proporciona un límite de predicción unilateral inferior calibrado aproximado al 95 % para el tiempo de vida del producto en campo. Además, un límite de predicción unilateral superior naive al 99.01 % proporciona un límite de predicción unilateral superior calibrado aproximado al 95 % para el tiempo de vida del producto en campo.

A continuación se presentan los límites de predicción unilaterales calibrados para el tiempo de vida del componente A en campo utilizando sólo la distribución de tiempo de vida dada por la ALT sin censura y asumiendo una tasa de uso constante:

Tabla 3.5: Límites de predicción unilaterales inferiores calibrados sin censura en la ALT

CP[PI(1 - α_{cl})] = 0.90 Sin Censura en la ALT			
Límite inferior naive Modelo	Límite inferior calibrado Modelo Método 1	Límite inferior naive ALT	Límite inferior calibrado ALT Método 1
[13.8146, + ∞)	[10.885443, + ∞)	[287.1873, + ∞)	[192.6669, + ∞)
CP[PI(1 - α_{cl})] = 0.95 Sin Censura en la ALT			
[10.19789, + ∞)	[7.496723, + ∞)	[219.1107, + ∞)	[129.7288, + ∞)

Tabla 3.6: Límites de predicción unilaterales superiores calibrados sin censura en la ALT

CP[PI(1 - α_{cu})] = 0.90 Sin Censura en la ALT			
Límite superior naive Modelo	Límite superior calibrado Modelo Método 1	Límite superior naive ALT	Límite superior calibrado ALT Método 1
(0, 120.4639]	(0, 154.5540]	(0, 1937.0533]	(0, 2899.603]
CP[PI(1 - α_{cu})] = 0.95 Sin Censura en la ALT			
(0, 164.47993]	(0, 223.1378]	(0, 2538.8860]	(0, 4232.557]

Para calcular un intervalo de predicción bilateral para el tiempo de vida del producto en campo al 90% sin censura, los límites de predicción unilaterales naive al 95% pueden ser combinados, como se muestra en la Tabla 3.7.

Tabla 3.7: Intervalos de predicción bilaterales calibrados sin censura

CP[PI(1 - α_c)] = 0.90 Sin Censura en la ALT			
Intervalo naive Modelo	Intervalo calibrado Modelo Método 1	Intervalo naive ALT	Intervalo calibrado ALT Método 1
[10.19789, 164.47993]	[7.496723, 223.1378]	[219.1107, 2538.8860]	[129.7288, 4232.557]

De acuerdo con estos resultados, con datos de vida sin censura y tomando la distribución de tiempo de vida dada sólo por la ALT, podemos estar 90 % seguros de que el componente A fallará entre la semana 129.73 y la semana 4232.56 de servicio, utilizando el Método 1 de calibración.

Si se comparan estos resultados con los obtenidos al utilizar el modelo de tasa de uso propuesto por Meeker *et al.* (2009) sin censura (ver Tabla 3.2 a la Tabla 3.7), las diferencias que se presentan son muy notables. Tanto los límites como los intervalos de predicción naive y calibrados usando sólo la distribución de tiempo de vida dada por la ALT y asumiendo 4 usos por semana como constante dentro de la población de consumidores, toman valores mucho más altos, por lo que podría estarse presentando una sobreestimación que desfasa los resultados deseados. Por lo tanto, si se hace la presunción de que el producto es usado con la misma intensidad por todos los clientes, lo cual no es coherente con lo que sucede en la realidad, podríamos obtener información errada en el momento de realizar la predicción estadística del tiempo de vida del producto en el campo. Por otro lado, también se puede afirmar que resulta mejor utilizar el modelo de Meeker *et al.* (2009) ya que para obtener intervalos de predicción calibrados al $(1 - \alpha)100\%$ se requieren intervalos naive al $(1 - \alpha_c)100\%$ de menor valor, por lo tanto, la aproximación dada por este tipo de intervalos es mejor que la que proporcionan los intervalos obtenidos a partir de la distribución del tiempo de vida para el componente A establecida en la ALT, lo cual puede evidenciarse en las curvas de calibración que se presentaron en esta Sección.

Conclusiones

A partir del estudio que se realizó en este trabajo de investigación con respecto a la predicción estadística para el modelo de tasa de uso para confiabilidad en campo propuesto por Meeker *et al.* (2009) se puede concluir lo siguiente:

- En la determinación del desempeño de un producto en manos del consumidor a partir de los resultados de pruebas de laboratorio, no siempre se logran identificar todas las variables asociadas, así como los diferentes modos de falla. Es por eso que resulta adecuado combinar esta información con la que se pueda tener del producto en campo, proveniente de las reclamaciones por garantías y de las tasas de uso del mismo, para establecer una mejor estimación de la distribución de tiempo de vida del producto cuando salga al mercado.
- Los intervalos de predicción naive son generalmente muy estrechos para la mayoría de los problemas, además tienen una probabilidad de cobertura que es inferior al valor nominal de $1 - \alpha$, por lo que no son exactos, esto debido a que ignoran la incertidumbre en $\hat{\theta}$ con respecto a θ . Esta deficiencia puede ser resuelta de alguna manera aplicando un procedimiento de calibración adecuado para dichos intervalos, así como se demostró en este trabajo al utilizar las técnicas propuestas por Escobar y Meeker (1999) para calibrar los límites de predicción naive para el tiempo de vida del componente A en campo.
- En cuanto a la estimación de los parámetros del modelo de tasa de uso para confiabilidad en campo, necesaria para realizar la predicción estadística del tiempo de vida, a medida que el tiempo de censura disminuye y por lo tanto se tiene menos tiempo para observar los ciclos hasta la falla de las unidades en la ALT y los tiempos de garantía para el producto en manos del cliente son menores, el parámetro $\hat{\eta}_C$ toma un valor menor, lo cual da una idea de que tiende a subestimarse. Este resultado se da porque en el momento de realizar la estimación se cuenta con menos información acerca de los tiempos de falla exactos de las unidades tanto en la ALT como en el campo. Por otro lado, la estimación dada por $\hat{\sigma}_C$, no presenta mucha variación entre los diferentes tiempos de censura.
- Un resultado importante que se observa en la predicción estadística para el modelo de tasa de uso, es que a medida que disminuyen los tiempos de censura, tanto los intervalos de predicción bilaterales naive como los calibrados tienden a ser más estrechos, esto puede explicarse por el hecho de que la estimación puntual dada por

$\hat{\eta}$ también toma un menor valor por lo que se produce una subestimación, lo cual ocurre análogamente con los límites de los intervalos, más específicamente con el límite superior.

- Con respecto a las metodologías de calibración para los límites de predicción naive propuestas por Escobar y Meeker (1999), se puede decir que los resultados arrojados por el Método 1 y Método 2, que son casos particulares de la calibración por simulación del proceso de muestreo/predicción y por el Método 3, donde la calibración se obtiene promediando las probabilidades de cobertura condicional, son muy similares, sin embargo, de acuerdo con la longitud de los intervalos calibrados obtenidos y teniendo en cuenta que tienen la misma probabilidad de cobertura que es 90 %, el método que proporciona los mejores resultados es el Método 2, ya que los intervalos son más cortos, además coincide con el hecho de que el algoritmo por medio del cual se lleva a cabo este procedimiento es computacionalmente más eficiente que el de las otras dos metodologías, esto debido a que los tiempos de las simulaciones son menores aún cuando los tamaños muestrales fueron mayores ($B = 50000$).
- Tanto los límites como los intervalos de predicción naive y calibrados usando sólo la distribución de tiempo de vida dada por la ALT y asumiendo la misma tasa de uso del producto dentro de la población de consumidores, toman valores muy altos con respecto a los obtenidos utilizando el modelo de tasa de uso propuesto por Meeker *et al.* (2009), por lo que podría estarse presentando una sobreestimación que desfasa los resultados deseados. Por lo tanto, si se hace la presunción de que el producto es usado con la misma intensidad por todos los clientes, lo cual no es coherente con lo que sucede en la realidad, podríamos obtener información errada en el momento de realizar la predicción estadística del tiempo de vida del producto en el campo.
- Se puede afirmar que resulta mejor utilizar el modelo de Meeker *et al.* (2009) ya que para obtener intervalos de predicción calibrados al $(1 - \alpha)100\%$ se requieren intervalos naive al $(1 - \alpha_c)100\%$ de menor valor, por lo tanto, la aproximación dada por este tipo de intervalos es mejor que la que proporcionan los intervalos obtenidos a partir de la distribución del tiempo de vida para el componente A establecida en la ALT con una tasa de uso igual para todos los consumidores.

Trabajo futuro

A partir de lo realizado en este trabajo, quedan aspectos por estudiar que se podrían explorar en futuros proyectos de investigación, entre los cuales se encuentran:

- Realizar todo el trabajo de predicción estadística para el modelo de tasa de uso para confiabilidad en campo propuesto por Meeker *et al.* (2009) bajo otras distribuciones diferentes a la lognormal.
- Implementar las metodologías propuestas por Escobar y Meeker (1999) para la construcción de intervalos y límites de predicción para el número de fallas futuras utilizando el modelo de tasa de uso para confiabilidad en campo.

Apéndice A

Rutinas implementadas en R

- Generación de los datos para el problema con el componente A

```
#####  
## Distribución estimada de ##  
## los ciclos de uso por semana ##  
#####  
  
#Generación de valores aleatorios lognormales  
k=rlnorm(100000,meanlog =1.6,sdlog=0.5)  
#Truncamiento de los valores en 20 ciclos  
t1=table(cut(k,breaks=0:20))  
#Conversión a frecuencias relativas  
t1=t1/sum(t1)  
#Distribución de las Tasas de Uso (histograma de la distribución  
#estimada de los ciclos por semana)  
barplot(t1, xlab="Frecuencia Relativa de los Usos por Semana",  
names.arg=c(1:20),cex.names=0.7)  
#Tabla con los valores de los ciclos y sus frecuencias  
#relativas respectivas  
tabla=as.data.frame(t1)  
tabla[1]=1:20  
  
#####  
## Distribución de los ciclos ##  
## hasta la falla en el ALT ##  
#####  
  
#Generación de ciclos hasta la falla a partir del ALT
```

```

ciclos1<-function(n,eta=200, sigma=1){
ciclos=rlnorm(n,meanlog=log(eta), sdlog =sigma)
return(ciclos)
}
resultado1=ciclos(200)

#####
## Generación de valores de la ##
## variable aleatoria discreta R #
#####

simdiscrete <- function(tabla, n.gen){
cum.probs <- cumsum(tabla[,2])
xprobs.sorted <- cbind(tabla,cum.probs)
# generate and loop
urandom <- runif(n.gen)
sim.vector <- rep(0,n.gen)
for(i in 1:n.gen)
sim.vector[i]<-min(xprobs.sorted[,1][which(cum.probs>= urandom[i])])
sim.vector
}
R <- simdiscrete(tabla,100)## tasas de uso

#####
### Distribución del tiempo ###
### de vida en campo ###
#####

#Generación de tiempos de falla a partir de la mezcla de distribuciones
tiempos1<-function(R1,eta=200, sigma=1){
tiempos=rlnorm(1,meanlog =log(eta/R1), sdlog =sigma)
return(tiempos)
}
resultado=sapply(R,tiempos)

```

- Estimación por máxima verosimilitud de los parámetros del modelo sin censura

```

tiempos<-as.vector(resultado)
R
Y=R*tiempos
ciclos=as.vector(resultado1)

#Función de log verocimilitud en ALT
alt<- function(theta,ciclos1){
totalt<-sum(-log(ciclos1)-log(theta[2]))+
dlnorm(ciclos1,meanlog= log(theta[1]),sdlog = theta[2],log = TRUE))

```

```

return(totalt)
}

#Función de log verocimilitud en campo
mixture<- function(theta,tiempos1,Y1){
logmix=sum(-log(tiempos1)-log(theta[2]))+
dlnorm(Y1,meanlog= log(theta[1]),sdlog = theta[2],log = TRUE))
return(logmix)
}

#Función de log verocimilitud (negativa) total
logtotalneg<-function(theta,tiempos1, Y1,ciclos1){
logtotal= -mixture(theta,tiempos1, Y1)-alt(theta,ciclos1)
return(logtotal)
}

#Optimización por Evolución Diferencial
library(DEoptim)
lower <- c(0,0)
upper <- c(500,500)
opt3<-DEoptim(logtotalneg,lower,upper,tiempos1=tiempos,Y1=Y,ciclos1=ciclos,
control=DEoptim.control(NP = 70,itermax = 400,trace = FALSE))

#Parámetros estimados sin censura
teta=as.vector(opt3$optim$bestmem)
teta

```

- Estimación por máxima verosimilitud de los parámetros del modelo con censura

```

##Cálculo de los percentiles de la mezcla con la función qnorMix
library(nor1mix)
eta1=200
sigma1=1
R=tabla$Var1
phi=tabla$Freq
mu=log(eta1/R)
sig2 = rep(sigma1^2,20)
mezcla=normix(mu, sig2, w = phi, name = NULL, long.name = FALSE)
quantiles=qnorMix(0.95, mezcla, lower.tail = TRUE)## Para el percentil 95
## De igual manera se hace para los percentiles 80 y 70
quantiles
exp(quantiles)
#####

##Cálculo de los percentiles de la mezcla con la función uniroot

```

```

Ft=function(t,tabla1=tabla,eta=eta1,sigma=sigma1){
f=sum(phi*pnorm((log(R*t)-log(eta))/sigma))
f
}
simul=function(t,p=0.95){## también para p=0.80 y p=0.70
r=Ft(t)-p
r
}
quantil=uniroot(simul,interval=c(0,15000),tol=0.0001)$root
quantil
#####

tiempos<-as.vector(resultado)
tiempos
R
a=257 #Criterio de censura para los tiempos de la mezcla percentil 0.95
## Para el percentil 95 a=257, percentil 80 a=104, percentil 70 a=73
delta<-ifelse(tiempos<a,1,0)
tiempos[delta==0]=a
Y=R*tiempos
Y
ciclos=as.vector(resultado1)
ciclos
b=quantile(ciclos,0.95)#Criterio de censura
para los ciclos en el ALT percentil 0.95
b=1042
## Para el percentil 95 b=1042, percentil 80 b=474, percentil 70 b=381
tao<-ifelse(ciclos<b,1,0)
ciclos[tao==0]=b

#Función de log verocimilitud en ALT
alt<- function(theta,ciclos1,tao1){
totalt=sum((-log(ciclos1)-log(theta[2]))+
dlnorm(ciclos1,meanlog=log(theta[1]),sdlog=theta[2],log=TRUE))*(tao1)
+(plnorm(ciclos1,meanlog=log(theta[1]),sdlog=theta[2],
lower.tail=FALSE,log.p=TRUE))*(1-tao1))
return(totalt)
}

#Función de log verocimilitud en campo
mixture<- function(theta,tiempos1,Y1,delta1){
logmix=sum((-log(tiempos1)-log(theta[2]))+
dlnorm(Y1,meanlog=log(theta[1]),sdlog=theta[2],log=TRUE))*(delta1)
+(plnorm(Y1,meanlog=log(theta[1]),sdlog=theta[2],
lower.tail=FALSE,log.p=TRUE))*(1-delta1))
return(logmix)
}

```

```

#Función de log verocimilitud (negativa) total
logtotalneg<-function(theta,tiempos1, Y1,ciclos1,delta1,tao1){
logtotal= -mixture(theta,tiempos1, Y1,delta1)-alt(theta,ciclos1,tao1)
return(logtotal)
}

#Optimización por Evolución Diferencial
library(DEoptim)
lower <- c(0,0)
upper <- c(500,500)
opt3<-DEoptim(logtotalneg,lower,upper,tiempos1=tiempos,Y1=Y,
ciclos1=ciclos,delta1=delta,tao1=tao,
control=DEoptim.control(NP=70,itermax=400,trace=FALSE))

#Parámetros estimados sin censura
teta=as.vector(opt3$optim$bestmem)
teta

```

- Intervalos y límites de predicción naive

```

## Intevalos de predicción bilaterales naive
Limite2=function(alfa, eta=200.3825271, sigma=0.6803828, tabla1=tabla){
phi=tabla1$Freq
R=tabla1$Var1
mu=log(eta/R)
sig2=rep(sigma^2,20)
mezcla=normMix(mu, sig2, w = phi)
quantiles1=qnormMix(alfa/2, mezcla, lower.tail = TRUE)
NLPB=exp(quantiles1)
quantiles2=qnormMix(1-(alfa/2), mezcla, lower.tail = TRUE)
NUPB=exp(quantiles2)
res=cbind(NLPB,NUPB)
return(res)
}
##Los valores para eta y sigma dentro de la función varían
de acuerdo a la censura

alfa=c(0.1,0.05)
limites2=sapply(alfa,Limite2)

## Límites de predicción unilaterales naive
Limite=function(nivel, eta=200.3825271, sigma=0.6803828, tabla1=tabla){
phi=tabla1$Freq
R=tabla1$Var1

```

```

mu=log(eta/R)
sig2=rep(sigma^2,20)
mezcla=normMix(mu, sig2, w = phi)
quantiles1=qnormMix(1-nivel, mezcla, lower.tail = TRUE)
NLPB=exp(quantiles1)
quantiles2=qnormMix(nivel, mezcla, lower.tail = TRUE)
NUPB=exp(quantiles2)
res=cbind(NLPB,NUPB)
return(res)
}
##Los valores para eta y sigma dentro de la función varían
de acuerdo a la censura

nivel=c(0.90,0.95)
limites=sapply(nivel,Limite)

```

- Calibración de los límites de predicción naive sin censura

```

##La primera parte es la misma para los tres Métodos:

##Generación de la muestra de los datos del ALT
ciclos1<-function(m,eta=200.3825271, sigma=0.6803828){
ciclos=rlnorm(m,meanlog=log(eta), sdlog =sigma)
return(ciclos)
}
# Función para simular valores de la distribución discreta de
##las tasas de uso
simdiscrete <- function(tab, n.gen){
cum.probs <- cumsum(tab[,2])
xprobs.sorted <- cbind(tab,cum.probs)
urandom <- runif(n.gen)
sim.vector <- rep(0,n.gen)
for(i in 1:n.gen)
sim.vector[i]<- min(xprobs.sorted[,1][which(cum.probs>=urandom[i])])
sim.vector
}
##Generación de la muestra de los datos en campo
tiempos1<-function(R1,eta=200.3825271, sigma=0.6803828){
tiempos=rlnorm(1,meanlog =log(eta/R1), sdlog =sigma)
return(tiempos)
}

## Estimación de los parámetros
mixture<- function(theta,tiempos1,Y1){
logmix=sum(-log(tiempos1)-log(theta[2])+log(dnorm(log(Y1),

```

```

mean=log(theta[1]), sd=theta[2]))
return(logmix)
}
alt<- function(theta,ciclos2){
totalt<-sum(-log(ciclos2)-log(theta[2])+log(dnorm(log(ciclos2),
mean=log(theta[1]), sd=theta[2])))
return(totalt)
}
logtotalneg<-function(theta,tiempos1,Y1,ciclos2){
logtotal= -mixture(theta,tiempos1,Y1)-alt(theta,ciclos2)
return(logtotal)
}

```

- Calibración de los límites de predicción naive sin censura: Método 1

```

nsimul=5000
##Función para estimar los j parámetros
CSSPP=function(tabla1=tabla,lower=c(0,0),upper=c(300,300)){
R<-simdiscrete(tabla1,100)
tiempos=as.vector(sapply(R,tiempos1))
Y=R*tiempos
ciclos=as.vector(ciclos1(200))
opt3<-DEoptim(logtotalneg, lower,upper, tiempos1=tiempos,
Y1=Y,ciclos2=ciclos,DEoptim.control(trace = FALSE))
teta=as.vector(opt3$optim$bestmem)
etaest=teta[1]
sigmaest=teta[2]
R2 <-simdiscrete(tabla1,1)
Tj=rlnorm(1, meanlog =log(200.3825271/R2), sdlog =0.6803828)
res=c(etaest,sigmaest,Tj)
return(res)
}

```

```

m=function(teta,nivel1,tabla1=tabla){
phi=tabla1$Freq
R=tabla1$Var1
mu=log(teta[1]/R)
sig2=rep(teta[2]^2,20)
mezcla=normMix(mu, sig2, w = phi)
quantiles1=qnormMix(1-nivel1, mezcla, lower.tail = TRUE)
NLPB=exp(quantiles1)
quantiles2=qnormMix(nivel1, mezcla, lower.tail = TRUE)
NUPB=exp(quantiles2)
comparacion1=ifelse(teta[3]>NLPB,1,0)
comparacion2=ifelse(teta[3]<NUPB,1,0)

```

```

res=cbind(comparacion1,comparacion2)
return(res)
}
CSSPP2=function(nivel2,nsim=nsimul){
teta1=matrix(replicate(nsim,CSSPP()),ncol=3,byrow=T)
resultado1=apply(teta1,1,m,nivel1=nivel2)
resultado2=apply(resultado1,1,mean)
resultado3=cbind(nivel2,t(resultado2))
colnames(resultado3)=c("nivel","CPI","CPS")
return(resultado3)
}

```

- Calibración de los límites de predicción naive sin censura: Método 2

```

##Función para estimar los j parámetros
CSSPP=function(tabla1=tabla,lower=c(0,0),upper=c(300,300)){
R1<-simdiscrete(tabla1,100)
tiempos=as.vector(sapply(R1,tiempos1))
Y=R1*tiempos
ciclos=as.vector(ciclos1(200))
opt3<-DEoptim(logtotalneg, lower, upper, tiempos1=tiempos,
Y1=Y, ciclos2=ciclos,DEoptim.control(trace = FALSE))
teta=as.vector(opt3$optim$bestmem)
etaest=teta[1]
sigmaest=teta[2]
R2 <-simdiscrete(tabla1,1)
Tj=rlnorm(1, meanlog =log(200.3825271/R2), sdlog =0.6803828)
phi=tabla1$Freq
R=tabla1$Var1
LPj=1-sum(phi*pnorm(log(Tj), mean=log(etaest/R), sd=sigmaest))
UPj=sum(phi*pnorm(log(Tj), mean=log(etaest/R), sd=sigmaest))
res=cbind(LPj,UPj)
return(res)
}

```

- Calibración de los límites de predicción naive sin censura: Método 3

```

nsimul=5000
##Función para estimar los j parámetros
CSSPP=function(tabla1=tabla,lower=c(0,0),upper=c(300,300)){
R<-simdiscrete(tabla1,100)
tiempos=as.vector(sapply(R,tiempos1))
Y=R*tiempos

```

```

ciclos=as.vector(ciclos1(200))
opt3<-DEoptim(logtotalneg, lower, upper, Y1=Y, tiempos1=tiempos,
ciclos2=ciclos, DEoptim.control(trace = FALSE))
teta=as.vector(opt3$optim$bestmem)
etaest=teta[1]
sigmaest=teta[2]
res=c(etaest,sigmaest)
return(res)
}

m=function(teta,nivel1,tabla1=tabla){
phi=tabla1$Freq
R=tabla1$Var1
mu=log(teta[1]/R)
sig2=rep(teta[2]^2,20)
mezcla=normMix(mu, sig2, w = phi)
quantiles1=qnormMix(nivel1, mezcla, lower.tail = FALSE)
NLPB=exp(quantiles1)
quantiles2=qnormMix(nivel1, mezcla, lower.tail = TRUE)
NUPB=exp(quantiles2)
LPj=1-sum(phi*pnorm(log(R*NLPB), mean=log(200.3825271), sd=0.6803828))
UPj=sum(phi*pnorm(log(R*NUPB), mean=log(200.3825271), sd=0.6803828))
res=cbind(LPj,UPj)
return(res)
}

CSSPP2=function(nivel2,nsim=nsimul){
teta1=matrix(replicate(nsim,CSSPP()),ncol=2,byrow=T)
resultado1=apply(teta1,1,m,nivel1=nivel2)
resultado2=apply(resultado1,1,mean)
resultado3=cbind(nivel2,t(resultado2))
colnames(resultado3)=c("nivel","CPI","CPS")
return(resultado3)
}

```

- Calibración de los límites de predicción naive con censura

```

##La primera parte es la misma para los tres Métodos:

##Generación de la muestra de los datos del ALT
ciclos1<-function(m,eta=198.7205186, sigma=0.6721745){
ciclos=rlnorm(m,meanlog=log(eta), sdlog =sigma)
return(ciclos)
}

```

```

# Función para simular valores de la distribución discreta de
##las tasas de uso
simdiscrete <- function(tab, n.gen){
  cum.probs <- cumsum(tab[,2])
  xprobs.sorted <- cbind(tab,cum.probs)
  urandom <- runif(n.gen)
  sim.vector <- rep(0,n.gen)
  for(i in 1:n.gen)
  sim.vector[i]<- min(xprobs.sorted[,1][which(cum.probs>= urandom[i])])
  sim.vector
}
##Generación de la muestra de los datos en campo
tiempos1<-function(R1,eta=198.7205186, sigma=0.6721745){
  tiempos=rlnorm(1,meanlog =log(eta/R1), sdlog =sigma)
  return(tiempos)
}

##Estimación de los parámetros
mixture<- function(theta,tiempos1, Y1,delta1){
  logmix=sum((-log(tiempos1)-log(theta[2])+dnorm(log(Y1),
  mean= log(theta[1]),sd= theta[2],
  log=TRUE))*(delta1)+(pnorm(log(Y1), mean= log(theta[1]),
  sd= theta[2], lower.tail = FALSE, log.p=TRUE))*(1-delta1))
  return(logmix)
}
alt<- function(theta,ciclos1,tao1){
  totalt=sum((-log(ciclos1)-log(theta[2])+dnorm(log(ciclos1),
  mean= log(theta[1]),sd= theta[2],
  log=TRUE))*(tao1)+(pnorm(log(ciclos1), mean= log(theta[1]),
  sd= theta[2], lower.tail = FALSE,log.p=TRUE))*(1-tao1))
  return(totalt)
}
logtotalneg<-function(theta,tiempos1, Y1,ciclos1,delta1,tao1){
  logtotal=-mixture(theta,tiempos1,Y1,delta1)
  -alt(theta,ciclos1,tao1)
  return(logtotal)
}

##Los valores para eta y sigma dentro de las funciones anteriores
varían de acuerdo a la censura

```

- Calibración de los límites de predicción naive con censura: Método 1

```

nsimul=5000
##Función para estimar los j parámetros

```

```

CSSPP=function(tabla1=tabla,lower=c(0,0),upper=c(300,300)){
R<-simdiscrete(tabla1,100)
tiempos=as.vector(sapply(R,tiempos1))
a=257## Criterio de Censura para los tiempos de la Mezcla
delta<-ifelse(tiempos<a,1,0)
tiempos[delta==0]=a
Y=R*tiempos
ciclos=as.vector(ciclos1(200))
b=1042# cuntil de la muestra original de los ciclos
tao<-ifelse(ciclos<b,1,0)
ciclos[tao==0]=b
opt3<-DEoptim(logtotalneg, lower, upper, tiempos1=tiempos,Y1=Y,
ciclos1=ciclos, delta1=delta, tao1=tao,DEoptim.control(trace = FALSE))
teta=as.vector(opt3$optim$bestmem)
etaest=teta[1]
sigmaest=teta[2]
R2 <-simdiscrete(tabla1,1)
Tj=rlnorm(1, meanlog =log(198.7205186/R2), sdlog =0.6721745)
res=c(etaest,sigmaest,Tj)
return(res)
}

##Los valores para eta, sigma, a y b dentro de las funciones anteriores
varían de acuerdo a la censura

m=function(teta,nivel1,tabla1=tabla){
phi=tabla1$Freq
R=tabla1$Var1
mu=log(teta[1]/R)
sig2=rep(teta[2]^2,20)
mezcla=norMix(mu, sig2, w = phi)
quantiles1=qnorMix(nivel1, mezcla, lower.tail = FALSE)
NLPB=exp(quantiles1)
quantiles2=qnorMix(nivel1, mezcla, lower.tail = TRUE)
NUPB=exp(quantiles2)
comparacion1=ifelse(teta[3]>NLPB,1,0)
comparacion2=ifelse(teta[3]<NUPB,1,0)
res=cbind(comparacion1,comparacion2)
return(res)
}

CSSPP2=function(nivel2,nsim=nsimul){
teta1=matrix(replicate(nsim,CSSPP()),ncol=3,byrow=T)
resultado1=apply(teta1,1,m,nivel1=nivel2)
resultado2=apply(resultado1,1,mean)
resultado3=cbind(nivel2,t(resultado2))
colnames(resultado3)=c("nivel","CPI","CPS")
return(resultado3)
}

```

```
}

```

- Calibración de los límites de predicción naive con censura: Método 2

```
##Función para estimar los j parámetros
CSSPP=function(tabla1=tabla,lower=c(0,0),upper=c(300,300)){
R<-simdiscrete(tabla1,100)
tiempos=as.vector(sapply(R,tiempos1))
a=257## Criterio de Censura para los tiempos de la Mezcla
delta<-ifelse(tiempos<a,1,0)
tiempos[delta==0]=a
Y=R*tiempos
ciclos=as.vector(ciclos1(200))
b=1042 # cuntil de la muestra original de los ciclos
tao<-ifelse(ciclos<b,1,0)
ciclos[tao==0]=b
opt3<-DEoptim(logtotalneg, lower, upper,tiempos1=tiempos, Y1=Y,
ciclos1=ciclos, delta1=delta, tao1=tao,DEoptim.control(trace = FALSE))
teta=as.vector(opt3$optim$bestmem)
etaest=teta[1]
sigmaest=teta[2]
R2 <-simdiscrete(tabla1,1)
Tj=rlnorm(1, meanlog =log(198.7205186/R2), sdlog =0.6721745)
phi=tabla1$Freq
R=tabla1$Var1
LPj=1-sum(phi*pnorm(log(Tj), mean=log(etaest/R), sd=sigmaest))
UPj=sum(phi*pnorm(log(Tj), mean=log(etaest/R), sd=sigmaest))
res=cbind(LPj,UPj)
return(res)
}
##Los valores para eta, sigma, a y b dentro de las funciones anteriores
varían de acuerdo a la censura

## Esta pareda final es la misma para el Método 2
## sin censura y con censura

nivel=matrix(seq(0.9,0.99,len=50))
# Distribución de cada v.a. P (50000 valores)
resultado=replicate(50000,CSSPP())
# Cálculo de los cuantiles para los diferentes valores de alpha
resultados=apply(resultado,2,quantile,probs =nivel)
## 1-alfa calibrados
resultados=cbind(nivel,resultados)
resultados

```

```

tabla=as.data.frame(t1)
tabla[1]=1:20
tabla

suav1=smooth.spline(resultados[,1:2])
suav2=smooth.spline(resultados[,c(1,3)])
#valores suavizados de los CPI
suav1$y
suav2$y

nivel=matrix(seq(0.9,0.99,len=50))
resultados=cbind(nivel,suav1$y,suav2$y)
alfainf_0.95=(resultados[,2][resultados[,1]<=0.9510000&
resultados[,1]>0.9490000])
alfasup_0.95=(resultados[,3][resultados[,1]<=0.9510000&
resultados[,1]>0.9490000])
alfainf_0.90=(resultados[,2][resultados[,1]<=0.9010000&
resultados[,1]>0.8990000])
alfasup_0.90=(resultados[,3][resultados[,1]<=0.9010000&
resultados[,1]>0.8990000])

```

- Calibración de los límites de predicción naive con censura: Método 3

```

nsimul=5000
##Función para estimar los j parámetros
CSSPP=function(tabla1=tabla,lower=c(0,0),upper=c(300,300)){
R<-simdiscrete(tabla1,100)
tiempos=as.vector(sapply(R,tiempos1))
a=257## Criterio de Censura para los tiempos de la Mezcla
delta<-ifelse(tiempos<a,1,0)
tiempos[delta==0]=a
Y=R*tiempos
ciclos=as.vector(ciclos1(200))
b=1042# cuntil de la muestra original de los ciclos
tao<-ifelse(ciclos<b,1,0)
ciclos[tao==0]=b
opt3<-DEoptim(logtotalneg, lower, upper,tiempos1=tiempos, Y1=Y,
ciclos1=ciclos, delta1=delta, tao1=tao,DEoptim.control(trace = FALSE))
teta=as.vector(opt3$optim$bestmem)
etaest=teta[1]
sigmaest=teta[2]
res=c(etaest,sigmaest)
return(res)
}
m=function(teta,nivel1){

```

```

phi=tabla$Freq
R=tabla$Var1
mu=log(teta[1]/R)
sig2=rep(teta[2]^2,20)
mezcla=normMix(mu, sig2, w = phi)
quantiles1=qnormMix(nivel1, mezcla, lower.tail = FALSE)
NLPB=exp(quantiles1)
quantiles2=qnormMix(nivel1, mezcla, lower.tail = TRUE)
NUPB=exp(quantiles2)
LPj=1-sum(phi*pnorm(log(R*NLPB), mean=log(198.7205186), sd=0.6721745))
UPj=sum(phi*pnorm(log(R*NUPB), mean=log(198.7205186), sd=0.6721745))
res=cbind(LPj,UPj)
return(res)
}

##Los valores para eta, sigma, a y b dentro de las funciones anteriores
varían de acuerdo a la censura

## Esta parte final es la misma para el Método 1 y Método 3
## sin censura y con censura

CSSPP2=function(nivel2,nsim=nsimul){
teta1=matrix(replicate(nsim,CSSPP()),ncol=2,byrow=T)
resultado1=apply(teta1,1,m,nivel1=nivel2)
resultado2=apply(resultado1,1,mean)
resultado3=cbind(nivel2,t(resultado2))
colnames(resultado3)=c("nivel","CPI","CPS")
return(resultado3)
}

#matriz que contiene los 1-alfa ordenados en una sola columna
nivel=matrix(seq(0.9,0.99,len=50))
## 1-alfa calibrados
resultados=t(apply(nivel,1,CSSPP2))
resultados

tabla=as.data.frame(t1)
tabla[1]=1:20
tabla

#valores suavizados de los CPI
suav1$y
suav2$y
#función para los alfas calibrados
calib=function(x,suav,CP){
f=predict(suav,x)$y-CP
f
}

```

```
#Interpolación para hallar los alfas calibrados
alfainf_0.90=uniroot(calib,interval=c(0,1),suav=suav1,CP=0.90)$root
alfasup_0.90=uniroot(calib,interval=c(0,1),suav=suav2,CP=0.90)$root
alfainf_0.95=uniroot(calib,interval=c(0,1),suav=suav1,CP=0.95)$root
alfasup_0.95=uniroot(calib,interval=c(0,1),suav=suav2,CP=0.95)$root
```

- Límites de predicción naive calibrados

```
## Límites de predicción calibrados
Limiteinfcal=function(alfainf,eta=200.3825271,
sigma=0.6803828, tabla1=tabla){
phi=tabla$Freq
R=tabla$Var1
mu=log(eta/R)
sig2=rep(sigma^2,20)
mezcla=normMix(mu, sig2, w = phi)
quantiles1=qnormMix(1-alfainf, mezcla, lower.tail = TRUE)
NLPBcal=exp(quantiles1)
res=NLPBcal
return(res)
}
Limitesupcal=function(alfasup, eta=200.3825271,
sigma=0.6803828, tabla1=tabla){
phi=tabla$Freq
R=tabla$Var1
mu=log(eta/R)
sig2=rep(sigma^2,20)
mezcla=normMix(mu, sig2, w = phi)
quantiles2=qnormMix(alfasup, mezcla, lower.tail = TRUE)
NUPBcal=exp(quantiles2)
res=NUPBcal
return(res)
}

##Los valores para eta y sigma dentro de la función varían
de acuerdo a la censura

alfainf=c(alfainf_0.90,alfainf_0.93,alfainf_0.95)
limiteinfcal=sapply(alfainf,Limiteinfcal)
alfasup=c(alfasup_0.90,alfainf_0.93,alfasup_0.95)
limitesupcal=sapply(alfasup,Limitesupcal)
```

- Gráficos curvas de calibración Método 1 y Método 3

```

#Límite inferior 0.90
plot(suav1,type='l',ylim=c(0.85,0.99),xlim=c(0.9,0.99),
ylab=expression(paste("CP",sep=" ","[" ,sep=" ", "PI",sep=" ",
(1-alpha[cal]),sep=" ",";"," ,sep=" ",hat(theta),sep=" ","]")),
cex.main=1, xlab=expression(1-alpha[cal]), lty=1, lwd=1)
#legend("topleft",legend="Límite Inferior",lty=1,col=1, lwd=1)
text(0.9085,0.90,expression(CPB==0.90),pos=3)
abline(h=0.90,col=4)
abline(v=alfainf_0.90,col=4)
text(0.907,0.938,expression(1-alpha==0.90))
text(0.9065,0.93,expression(1-alpha[cal_inf]))
text(0.922,0.93,labels=paste("=",sep=" ",round(alfainf_0.90,5)))
text(0.9255,0.946,"Número de muestras simuladas=5000")

#Límite superior 0.90
plot(suav2,type='l',ylim=c(0.85,0.99),xlim=c(0.9,0.99),col=2, lty=2,
lwd=1,ylab=expression(paste("CP",sep=" ","[" ,sep=" ", "PI",sep=" ",
(1-alpha[cal]),sep=" ",";"," ,sep=" ",hat(theta),sep=" ","]")),
cex.main=1, xlab=expression(1-alpha[cal]))
#legend("topleft",legend="Límite Superior",lty=2,col=2, lwd=1)
text(0.9085,0.90,expression(CPB==0.90),pos=3)
abline(h=0.90,col=4)
abline(v=alfasup_0.90,col=4)
text(0.906,0.938,expression(1-alpha==0.90))
text(0.9065,0.93,expression(1-alpha[cal_sup]))
text(0.922,0.93,labels=paste("=",sep=" ",round(alfasup_0.90,5)))
text(0.924,0.946,"Número de muestras simuladas=5000")

#Límite inferior 0.95
par(mar=c(4.5,5,2,2))
plot(suav1,type='l',ylim=c(0.85,0.99),xlim=c(0.9,0.99),
ylab=expression(paste("CP",sep=" ","[" ,sep=" ", "PI",sep=" ",
(1-alpha[cal]),sep=" ",";"," ,sep=" ",hat(theta),sep=" ","]")),
cex.main=1, xlab=expression(1-alpha[cal]), lty=1, lwd=1)
#legend("topleft",legend="Límite Inferior",lty=1,col=1, lwd=1)
text(0.9085,0.95,expression(CPB==0.95),pos=3)
abline(h=0.95,col=4)
abline(v=alfainf_0.95,col=4)
text(0.9092,0.938,expression(1-alpha==0.95))
text(0.908,0.93,expression(1-alpha[cal_inf]))
text(0.922,0.93,labels=paste("=",sep=" ",round(alfainf_0.95,5)))
text(0.9265,0.946,"Número de muestras simuladas=5000")

#Límite superior 0.95
plot(suav2,type='l',ylim=c(0.85,0.99),xlim=c(0.9,0.99),col=2, lty=2,
lwd=1,ylab=expression(paste("CP",sep=" ","[" ,sep=" ", "PI",sep=" ",

```

```
(1-alpha[cal]),sep=" ";",sep=" ",hat(theta),sep=" ","]")),
cex.main=1, xlab=expression(1-alpha[cal]))
#legend("topleft",legend="Límite Superior",lty=2,col=2, lwd=1)
text(0.9085,0.95,expression(CPB==0.95),pos=3)
abline(h=0.95,col=4)
abline(v=alfasup_0.95,col=4)
text(0.9092,0.938,expression(1-alpha==0.95))
text(0.908,0.93,expression(1-alpha[cal_sup]))
text(0.922,0.93,labels=paste("=",sep=" ",round(alfasup_0.95,5)))
text(0.9265,0.946,"Número de muestras simuladas=5000")
```

- Gráficos curvas de calibración Método 2

```
#Límite inferior 0.90
plot(suav1$y,nivel,type='l',ylim=c(0.9,0.99),xlim=c(0.94,1),
ylab=expression(paste("CP",sep=" ","[" ,sep=" ", "PI",sep=" ",
(1-alpha[cal]),sep=" ";",sep=" ",hat(theta),sep=" ","]")),
cex.main=1, xlab=expression(1-alpha[cal]), lty=1, lwd=1)
#legend("topleft",legend="Límite Inferior",lty=1,col=1, lwd=1)
text(0.955,0.90,expression(CPB==0.90),pos=3)
abline(h=0.90,col=4)
abline(v=alfainf_0.90,col=4)
text(0.9523,0.96,expression(1-alpha==0.90))
text(0.952,0.955,expression(1-alpha[cal_inf]))
text(0.963,0.955,labels=paste("=",sep=" ",round(alfainf_0.90,5)))
text(0.964,0.966,"Número de muestras simuladas=50000")
```

```
#Límite superior 0.90
plot(suav2$y,nivel,type='l',ylim=c(0.9,0.99),xlim=c(0.94,1),
ylab=expression(paste("CP",sep=" ","[" ,sep=" ", "PI",sep=" ",
(1-alpha[cal]),sep=" ";",sep=" ",hat(theta),sep=" ","]")),
cex.main=1, xlab=expression(1-alpha[cal]),col=2, lty=2)
#legend("topleft",legend="Límite Superior",lty=2,col=2, lwd=1)
text(0.955,0.90,expression(CPB==0.90),pos=3)
abline(h=0.90,col=4)
abline(v=alfasup_0.90,col=4)
text(0.952,0.96,expression(1-alpha==0.90))
text(0.952,0.955,expression(1-alpha[cal_sup]))
text(0.962,0.955,labels=paste("=",sep=" ",round(alfasup_0.90,5)))
text(0.962,0.966,"Número de muestras simuladas=50000")
```

```
#Límite inferior 0.95
par(mar=c(4.5,5,2,2))
plot(suav1$y,nivel,type='l',ylim=c(0.9,1),xlim=c(0.94,1),
ylab=expression(paste("CP",sep=" ","[" ,sep=" ", "PI",sep=" ",
```

```
(1-alpha[cal]),sep=" ",";",sep=" ",hat(theta),sep=" ","]")),
cex.main=1, xlab=expression(1-alpha[cal]), lty=1, lwd=1)
#legend("topleft",legend="Límite Inferior",lty=1,col=1, lwd=1)
text(0.945,0.951,expression(CPB==0.95),pos=3)
abline(h=0.95,col=4)
abline(v=alfainf_0.95,col=4)
text(0.9445,0.97,expression(1-alpha==0.95))
text(0.9436,0.965,expression(1-alpha[cal_inf]))
text(0.9536,0.965,labels=paste("=",sep=" ",round(alfainf_0.95,5)))
text(0.956,0.976,"Número de muestras simuladas=50000")
```

```
#Límite superior 0.95
plot(suav2$y,nivel,type='l',ylim=c(0.9,0.99),xlim=c(0.94,1),
ylab=expression(paste("CP",sep=" ","[" ,sep=" ", "PI",sep=" ",
(1-alpha[cal]),sep=" ",";",sep=" ",hat(theta),sep=" ","]")),
cex.main=1, xlab=expression(1-alpha[cal]),col=2, lty=2)
#legend("topleft",legend="Límite Superior",lty=2,col=2, lwd=1)
text(0.945,0.951,expression(CPB==0.95),pos=3)
abline(h=0.95,col=4)
abline(v=alfasup_0.95,col=4)
text(0.9445,0.97,expression(1-alpha==0.95))
text(0.9438,0.965,expression(1-alpha[cal_sup]))
text(0.9536,0.965,labels=paste("=",sep=" ",round(alfasup_0.95,5)))
text(0.956,0.976,"Número de muestras simuladas=50000")
```

- Estimación por máxima verosimilitud de los parámetros del modelo con información sólo de la ALT

```
R=4
ciclos2<-function(n,eta=200, sigma=1,R1){
ciclos=rlnorm(n,meanlog=log(eta*R1), sdlog =sigma)
return(ciclos)
}
resultado2=ciclos2(200,R1=R)
ciclos=as.vector(resultado2)
R=4
ciclos
K=ciclos/R
K

#Logverosimilitud para el ALT usando la normal
alt<- function(theta,ciclos1,K1){
totalt<-sum(-log(ciclos1)-log(theta[2])+log(dnorm(log(K1),
mean=log(theta[1]), sd=theta[2])))
```

```

return(totalt)
}
#Logverosimilitud Total negativa
logtotalneg<-function(theta,ciclos1,K1){
logtotal=-alt(theta,ciclos1,K1)
return(logtotal)
}
#Optimización con algoritmos evolutivos
library(DEoptim)
lower <- c(0,0)
upper <- c(500,500)
opt3<-DEoptim(logtotalneg, lower, upper, ciclos1=ciclos, K1=K,
control=DEoptim.control(NP = 70, itermax = 400, trace = FALSE))

## Teta sin censura
teta=as.vector(opt3$optim$bestmem)
teta

```

- Intervalos y límites de predicción naive con información sólo de la ALT

```

## Intevalos de predicción bilaterales
Limite2=function(alfa,eta=186.4633196 , sigma=0.7447178,R1=4){
NLPB=exp(log(eta*R1)+qnorm(alfa/2)*sigma)
NUPB=exp(log(eta*R1)+qnorm(1-(alfa/2))*sigma)
res=cbind(NLPB,NUPB)
return(res)
}
alfa=c(0.1,0.05)
limites2=sapply(alfa,Limite2)

## Límites de predicción unilaterales
Limite=function(nivel,eta=186.4633196,sigma=0.7447178,R1=4){
NLPB=exp(log(eta*R1)+(qnorm(1-nivel))*sigma)
NUPB=exp(log(eta*R1)+(qnorm(nivel))*sigma)
res=cbind(NLPB,NUPB)
return(res)
}
nivel=c(0.90,0.95)
limites=sapply(nivel,Limite)

```

- Calibración de los límites de predicción naive con información sólo de la ALT:Método1

```

library(DEoptim)
nsimul=5000
ciclos1<-function(m,eta=186.4633196,sigma=0.7447178,R1=4){
  ciclos=rlnorm(m,meanlog=log(eta*R1),sdlog =sigma)
  return(ciclos)
}
alt<- function(theta,ciclos2,K1){
  totalt<-sum(-log(ciclos2)-log(theta[2])+log(dnorm(log(K1),
  mean=log(theta[1]), sd=theta[2])))
  return(totalt)
}
logtotalneg<-function(theta,ciclos2,K1){
  logtotal= -alt(theta,ciclos2,K1)
  return(logtotal)
}

##Función para estimar los j parámetros
CSSPP=function(lower=c(0,0),upper=c(300,300)){
  R=4
  ciclos=as.vector(ciclos1(200))
  K=ciclos/R
  opt3<-DEoptim(logtotalneg, lower,upper,K1=K,
  ciclos2=ciclos,DEoptim.control(trace = FALSE))
  teta=as.vector(opt3$optim$bestmem)
  etaest=teta[1]
  sigmaest=teta[2]
  Tj=rlnorm(1, meanlog =log(186.4633196*R), sdlog =0.7447178)
  res=c(etaest,sigmaest,Tj)
  return(res)
}
m=function(teta,nivel1,tabla1=tabla,R=4){
  NLPB=exp(log(teta[1]*R)+(qnorm(1-nivel1))*teta[2])
  NUPB=exp(log(teta[1]*R)+(qnorm(nivel1))*teta[2])
  comparacion1=ifelse(teta[3]>NLPB,1,0)
  comparacion2=ifelse(teta[3]<NUPB,1,0)
  res=cbind(comparacion1,comparacion2)
  return(res)
}
CSSPP2=function(nivel2,nsim=nsimul){
  teta1=matrix(replicate(nsim,CSSPP()),ncol=3,byrow=T)
  resultado1=apply(teta1,1,m,nivel1=nivel2)
  resultado2=apply(resultado1,1,mean)
  resultado3=cbind(nivel2,t(resultado2))
  colnames(resultado3)=c("nivel","CPI","CPS")
  return(resultado3)
}

```

```

#matriz que contiene los 1-alfa ordenados en una sólo columna
nivel=matrix(seq(0.9,0.99,len=50))
## 1-alfa calibrados
resultados=t(apply(nivel,1,CSSPP2))
resultados

tabla=as.data.frame(t1)
tabla[1]=1:20
tabla

#valores suavizados de los CPI
suav1$y
suav2$y
#función para los alfas calibrados
calib=function(x,suav,CP){
f=predict(suav,x)$y-CP
f
}

#Interpolación para hallar los alfas calibrados
alfainf_0.90=uniroot(calib,interval=c(0,1),suav=suav1,CP=0.90)$root
alfasup_0.90=uniroot(calib,interval=c(0,1),suav=suav2,CP=0.90)$root
alfainf_0.95=uniroot(calib,interval=c(0,1),suav=suav1,CP=0.95)$root
alfasup_0.95=uniroot(calib,interval=c(0,1),suav=suav2,CP=0.95)$root

```

- Límites de predicción naive calibrados con información sólo de la ALT

```

Limiteinfcal=function(alfainf,eta=186.4633196,sigma=0.7447178,R1=4){
NLPBcal=exp(log(eta*R1)+(qnorm(1-alfainf))*sigma)
res=NLPBcal
return(res)
}
Limitesupcal=function(alfasup, eta=186.4633196,sigma=0.7447178,R1=4){
NUPBcal=exp(log(eta*R1)+(qnorm(alfasup))*sigma)
res=NUPBcal
return(res)
}
alfainf=c(alfainf_0.90,alfainf_0.95)
limiteinfcal=sapply(alfainf,Limiteinfcal)
alfasup=c(alfasup_0.90,alfasup_0.95)
limitesupcal=sapply(alfasup,Limitesupcal)

```

Bibliografía

- R Organización: *R-project*. Vienna, Austria, 2011 (URL: <http://www.R-project.org>)
- Antle, C. E. and F. Rademaker: An upper confidence limit on the maximum of m future observations from a type I extreme value distribution. *Biometrika*, **59** 1972, 475–477
- Atwood, C. L.: Approximate tolerance intervals, based on maximum likelihood estimates. *Journal of the American Statistical Association*, **79** 1984, 459–465
- Barndorff-Nielsen, O. and D. R. Cox: Prediction and asymptotics. *Bernoulli*, **2** 1996, 319–340
- Beran, R.: Calibrating prediction regions. *Journal of the American Statistical Association*, **85** 1990 Número 411, 715–723
- Beyer, H.G and H.P Schwefel: Evolution strategies A comprehensive introduction. *Natural Computing*, 2002, 3–52
- Cox, D. R.: Prediction intervals and empirical Bayes confidence intervals. En J. Gani, director: *Perspectives in Probability and Statistics*. London: Academic Press, 1975, 47–55
- Durbán, M.: An introduction to smoothing with penalties: P-splines. *Boletín de Estadística e Investigación Operativa*, **25** 2009 Número 3, 195–205
- Engelhardt, M. and L. J. Bain: Prediction limits and two-sample problems with complete or censored weibull data. *Technometrics*, **21** 1979, 233–237
- Escobar, L.A. and W. Meeker: Statistical prediction based on censored life data. *Technometrics*, **41** 1999 Número 2, 113–124
- Escobar, L.A. *et al.*: Confiabilidad: historia, estado del arte y desafíos futuros. *DYNA*, **70** 2003 Número 140, 5–21
- Fredette, M.: *Prediction of recurrent events*. Tesis Doctoral, University of Waterloo, Waterloo, Ontario, Canada, 2004
- Kim, C.S. *et al.*: Warranty cost analysis with heterogenous usage intensity. *International Transactions in Operational Research*, 2001 Número 8, 337–347

- Komaki, F.: On asymptotic properties of predictive distributions. *Biometrika*, **83** 1996, 299–313
- Lawless, J. F. and M. Fredette: Frequentist prediction intervals and predictive distributions. *Biometrika*, **92** Sep. 2005 Número 3, 529–542
- Manna, D. K. *et al.*: A use-rate based failure model for two-dimensional warranty. *Computers and Industrial Engineering*, **52** 2007, 229–240
- Mächler, M.: *nor1mix: Normal (1-d) Mixture Models (S3 Classes and Methods)*. **1.1-3** 2011 (URL: <http://cran.r-project.org/>)
- McLachlan, G. and D. Peel: *Finite mixture models*. New York: John Wiley and Sons, Inc., 2000
- Mee, R. and D. Kushary: Prediction limits for the weibull distribution utilizing simulation. *Computational Statistics and Data Analysis*, **17** 1994, 327–336
- Meeker, W. and L.A. Escobar: *Statistical methods for reliability data*. New York: Wiley, 1998
- Meeker, W. *et al.*: Using accelerated life tests results to predict product field reliability. *Technometrics*, **51** 2009, 146–161
- Morgan, B. J. T.: *Elements of simulation*. Chapman and Hall, 1984
- Nelson, W.: *Accelerated testing: statistical models, test plans and data analyses*. New York: Wiley, 1990
- Nelson, W.: Weibull prediction of a future number of failures. paper presented at the 1995 Joint Statistical Meetings, Toronto, 1995, August 13–17
- Nelson, W.: Prediction of field reliability of Units, each under differing dynamic stresses, from accelerated test data. in *Handbook of Statistics*, **20** 2001, eds. N. Balakrishnan and C. R. Rao
- Nelson, W. and J. Schmee: Prediction limits for the last failure of a (log) normal sample from early failures. *IEEE Transactions on Reliability*, **R-30** 1981, 461–463
- Titterton, D.M. *et al.*: *Statistical analysis of finite mixture distributions*. New York: John Wiley and Sons, Inc., 1985
- Vidoni, P.: A note on modified estimative prediction limits and distributions. *Biometrika*, **85** 1995, 949–953
- Yang, G. and Z. Zaghati: Accelerated life tests at higher usage rates: a case study. in the *Proceedings of the Reliability and Maintainability Symposium*, Philadelphia: IEEE, 2006, 313–317