

## INTERVALOS DE PREDICCIÓN PARA EL TIEMPO DE VIDA DE PRODUCTOS EN USO

### PREDICTION INTERVALS FOR THE LIFETIME OF PRODUCTS IN FIELD

ALEJANDRA ESTEFANÍA PATIÑO HOYOS<sup>1</sup> NELFI GERTRUDIS GONZÁLEZ ÁLVAREZ<sup>2</sup>

Recibido 13-11-12, aceptado 14-12-12, versión final 26-12-12.

Artículo Investigación

**RESUMEN:** En la actualidad 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 para obtener 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 estas pruebas de laboratorio con los datos de campo a través de la formulación de un modelo estadístico apropiado que permita establecer la relación entre ambos tipos de datos y su aprovechamiento en la determinación y estimación del modelo de vida 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 uso 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, con un modelo de tasa de uso del producto en el campo.

**PALABRAS CLAVE:** Confiabilidad, intervalos de predicción, modelo de tasa de uso, pruebas de vida acelerada.

**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 statistical model to establish the relationship between both types of data and use them in the identification and estimation of the model of 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 with a use rate model of product in the field.

**KEYWORDS:** Accelerated life tests, prediction intervals, reliability, use rate model.

---

<sup>1</sup>Estudiante de Maestría en Ciencias-Estadística, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín. [aepatino@unal.edu.co](mailto:aepatino@unal.edu.co)

<sup>2</sup>Profesora Asociada, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín. [ngonzale@unal.edu.co](mailto:ngonzale@unal.edu.co)

## 1 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 el 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 mismo en su campo de acción.

El problema a abordar en este artículo, es la predicción estadística del tiempo de vida de productos en uso 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.

## 2 MODELO DE TASA DE USO PARA CONFIABILIDAD EN EL CAMPO

Para los casos en los cuales no está disponible la información que permita caracterizar en el laboratorio las condiciones de uso del producto en campo, pero tanto los datos de pruebas de vida (ALT) 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 vida para un componente futuro o para un producto operando bajo las mismas condiciones de uso.

Dado que el objetivo de este artículo es la predicción estadística del tiempo de vida de productos en uso con base en el modelo de tasa de uso para confiabilidad en campo propuesto por Meeker et al. (2009), se partirá de un problema real con el componente A que ellos mismos plantean en su artículo, el cual se replicará vía simulación usando el paquete estadístico R (2011) 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 llamado componente A. Se realizó un cambio en el diseño con el cual se 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.

Para la mayoría de los modos de falla 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. Cualquiera que sea la definición de uso para un producto en particular, ésta debe depender del mecanismo de falla, es por eso que para el

componente A el tiempo se registró en los ciclos de uso del mismo.

Los ingenieros responsables de la confiabilidad de este producto proporcionaron la siguiente información (Meeker et al., 2009):

- 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, al 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 se muestra el histograma de la distribución estimada de los ciclos de uso (frecuencia 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.

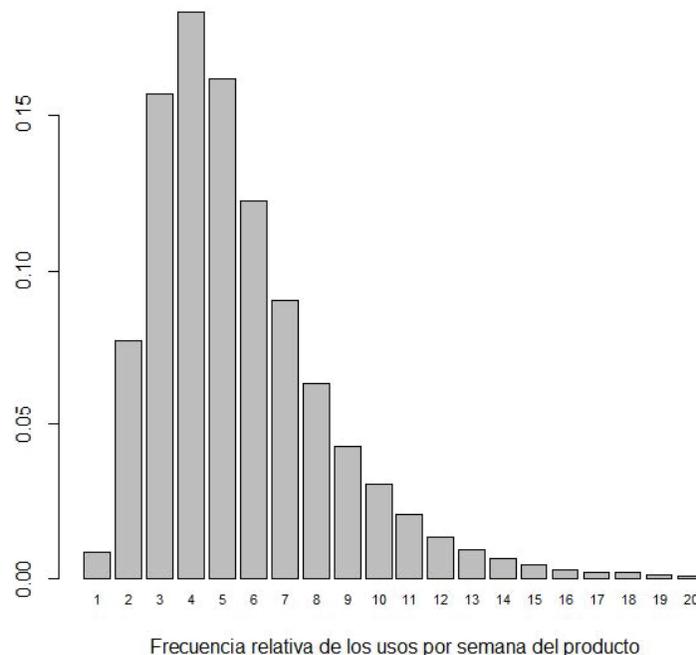


Figura 1: Distribución de la tasa de uso del componente A en los hogares

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. 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:

$$F_C(c) = \Pr(C \leq c) = \Phi_{nor} \left[ \frac{\log(c/\eta_C)}{\sigma_C} \right] = \Phi_{nor} \left[ \frac{\log(c) - \log(\eta_C)}{\sigma_C} \right]. \quad (1)$$

Donde  $\Phi$  es la función de densidad acumulada (cdf) normal,  $\eta_C$  es el parámetro de localización, la mediana en este caso, dado en unidades de los ciclos de la prueba y  $\sigma_C$  es el parámetro de forma. Entonces, el  $\log(C)$  se distribuye normal con parámetros  $\log(\eta_C)$  y  $\sigma_C^2$ , lo cual se denota

$$\log(C) \sim \mathbf{Normal}(\log(\eta_C), \sigma_C^2). \quad (2)$$

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

$$\mathbf{R} = (R_1, R_2, \dots, R_k), \quad (3)$$

y con probabilidades correspondientes:

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k), \text{ donde } \sum_{i=1}^k \pi_i = 1. \quad (4)$$

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 el campo, entonces la fracción de falla en el tiempo real  $t$  en unidades de las semanas de servicio en la población del producto, está descrita por una distribución de mezcla finita, que para este caso, debido a que los datos son simulados (ver Figura 1), se asume que es la siguiente mezcla finita de lognormales:

$$F(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \Pr(T \leq t) = \sum_{i=1}^k \pi_i \Phi_{nor} \left[ \frac{\log(R_i \times t/\eta_C)}{\sigma_C} \right]. \quad (5)$$

Resolviendo el logaritmo, puede reescribirse de la forma

$$F(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \Pr(T \leq t) = \sum_{i=1}^k \pi_i \Phi_{nor} \left[ \frac{\log(t) - \log(\eta_C/R_i)}{\sigma_C} \right]. \quad (6)$$

Por lo tanto, el  $\log(T)|\mathbf{R} = R_i$  se distribuye normal con parámetros  $\log(\eta_C/R_i)$  y  $\sigma_C^2$ , lo cual se denota

$$(\log(T)|\mathbf{R} = R_i) \sim \mathbf{Normal}(\log(\eta_C/R_i), \sigma_C^2). \quad (7)$$

Según Meeker et al. (2009), también es posible usar una distribución de mezcla continua para describir el 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.

Expuestas las características específicas del modelo de tasa de uso para el cual se pretenden construir los intervalos de predicción, es necesario presentar lo concerniente a la predicción estadística. Por lo tanto, en la siguiente sección se desarrollan los conceptos relacionados con este tema.

### 3 PREDICCIÓN ESTADÍSTICA

Las definiciones presentadas en esta Sección son tomadas de Meeker y Escobar (1998) y Escobar y Meeker (1999, cap. 12, pág. 289-304). 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, llamados también tiempos de vida, usando datos de una muestra previa. El método general usa calibración basada en simulación.

#### 3.1 Intervalos de predicción para tiempos de vida

Asumiendo que los tiempos de vida tienen una distribución de probabilidad continua completamente especificada  $F(t; \theta)$  y una función de densidad de probabilidad (pdf)  $f(t; \theta)$ , un intervalo de predicción PI (por sus siglas en inglés) exacto al  $(1 - \alpha)100\%$  para una observación futura  $T$  de  $F(t; \theta)$ , donde  $\alpha \in (0, 1)$  y  $\theta$  es un vector de parámetros que generalmente es desconocido por lo que puede ser estimado a partir de la muestra de datos disponibles, está dado por:

$$\text{PI}(1 - \alpha) = [T, \tilde{T}] = [t_{\alpha/2}, t_{1-\alpha/2}], \quad (8)$$

donde  $t_p$  es el cuantil  $p$  de  $F(t; \theta)$ . 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  $t_p$ .

Entonces, la probabilidad de cobertura del intervalo en (8) está dada por:

$$\begin{aligned} \Pr[T \in \text{PI}(1 - \alpha); \theta] &= \Pr[\underline{T} \leq T \leq \tilde{T}; \theta] \\ &= \Pr(t_{\alpha/2} \leq T \leq t_{1-\alpha/2}; \theta) \\ &= 1 - \alpha. \end{aligned} \quad (9)$$

#### 3.2 Probabilidad de cobertura para procedimientos de intervalos de predicción estadística para tiempos de vida (Caso en el que $\theta$ es estimado)

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

Existen dos tipos de probabilidades de cobertura (CP):

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

$$\begin{aligned} \text{CP}(\text{PI}(1 - \alpha)|\hat{\theta}; \theta) &= \Pr(T \leq T \leq \tilde{T}|\hat{\theta}; \theta) \\ &= F(\tilde{T}; \theta) - F(T; \theta). \end{aligned} \quad (10)$$

Esta probabilidad condicional es desconocida porque  $F(t; \theta)$  depende de  $\theta$  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:

$$\begin{aligned} \text{CP}(\text{PI}(1 - \alpha); \theta) &= \Pr(T \leq T \leq \tilde{T}; \theta) \\ &= E_{\hat{\theta}} \left\{ \text{CP}(\text{PI}(1 - \alpha)|\hat{\theta}; \theta) \right\}, \end{aligned} \quad (11)$$

donde el valor esperado es con respecto a  $\hat{\theta}$  aleatorio. Debido a que (11) 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  $\theta$ , se dice que el procedimiento PI es “exacto”. Cuando  $\text{CP}(\text{PI}(1 - \alpha); \theta) \neq 1 - \alpha$  no depende de  $\theta$ , generalmente es posible encontrar un procedimiento PI modificado que sea “exacto”. Cuando  $\text{CP}(\text{PI}(1 - \alpha); \theta)$  depende de un  $\theta$  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.

### 3.3 El método naive 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  $\theta$  en (8), dando

$$\text{PI}(1 - \alpha) = [T, \tilde{T}] = [\hat{t}_{\alpha/2}, \hat{t}_{1-\alpha/2}], \quad (12)$$

donde  $\hat{t}_p = t_p(\hat{\theta})$  es el MLE para el cuantil  $p$  de  $T$ . 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  $(1 - \alpha)$ .

### 3.4 Calibración de límites de predicción estadística naive

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  $CP[PI(1 - \alpha_c); \theta]$  en  $\hat{\theta}$  y encontrar un valor de calibración  $(1 - \alpha_c)$ , tal que, para un límite de predicción unilateral inferior para  $T$ :

$$\begin{aligned} CP[PI(1 - \alpha_{cl}); \hat{\theta}] &= \Pr(\underline{T} \leq T \leq \infty; \hat{\theta}) \\ &= \Pr(\hat{t}_{\alpha_{cl}} \leq T \leq \infty; \hat{\theta}) = 1 - \alpha, \end{aligned} \quad (13)$$

y para un límite de predicción unilateral superior para  $T$ :

$$\begin{aligned} CP[PI(1 - \alpha_{cu}); \hat{\theta}] &= \Pr(0 \leq T \leq \tilde{T}; \hat{\theta}) \\ &= \Pr(0 \leq T \leq \hat{t}_{1-\alpha_{cu}}; \hat{\theta}) = 1 - \alpha. \end{aligned} \quad (14)$$

En problemas en los cuales  $CP[PI(1 - \alpha_c); \theta]$  no depende de  $\theta$ , el procedimiento de calibración proporciona un intervalo de predicción exacto.

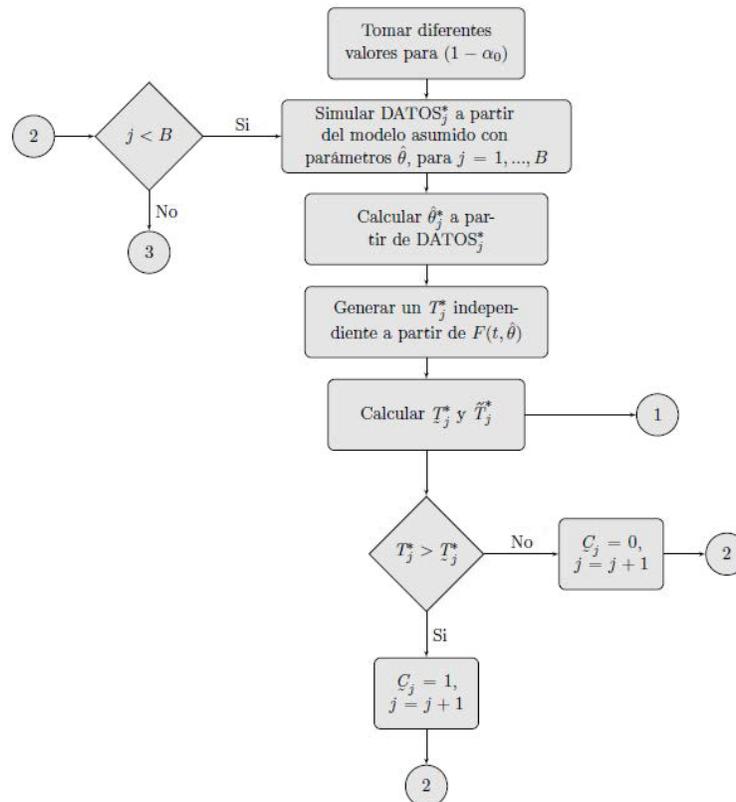


Figura 2: Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 1)

### 3.4.1 Método 1: Calibración por simulación del proceso de muestreo/predicción

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  $B$ . Conceptualmente, para obtener las curvas de calibración para los límites de predicción inferior y superior, la función  $CP[PI(1 - \alpha_c); \hat{\theta}]$  en (13) y (14) puede ser evaluada como lo indica el algoritmo presentado en las figuras 2 y 3.

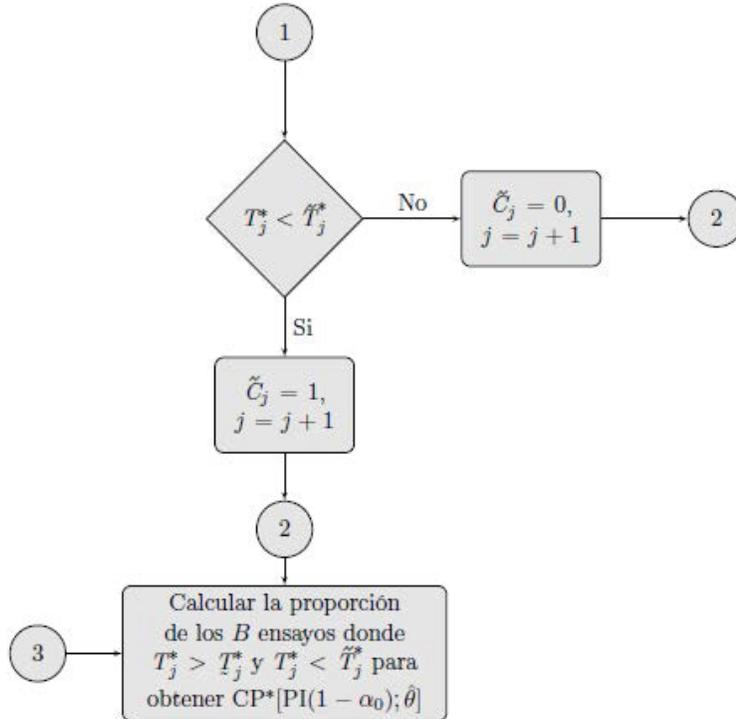


Figura 3: Continuación Figura 2: Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 1)

### 3.4.2 Método 2:

Operacionalmente, la función  $CP[PI(1 - \alpha_c); \hat{\theta}]$  completa en (13) y (14) puede ser evaluada más directamente, pero equivalentemente, usando el procedimiento escrito en el algoritmo presentando en la figura 4.

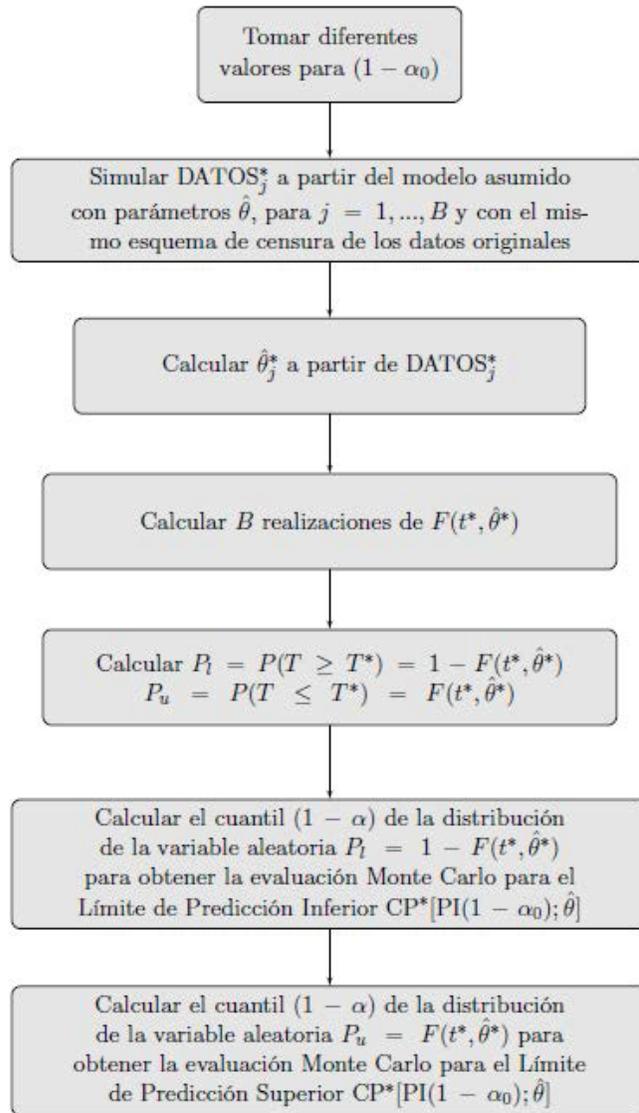


Figura 4: Algoritmo para la calibración de los límites de predicción inferior y superior por simulación del proceso de muestreo/predicción (Método 2)

### 3.4.3 Método 3: Calibración promediando las probabilidades de cobertura condicional

Una metodología alternativa para obtener las curvas de calibración en (13) y (14) es simular las probabilidades de cobertura condicional como las dadas en (10) y promediarlas para estimar el valor esperado en (11) (ver Mee y Kushary, 1994). El procedimiento a seguir se describe en la figura 5.

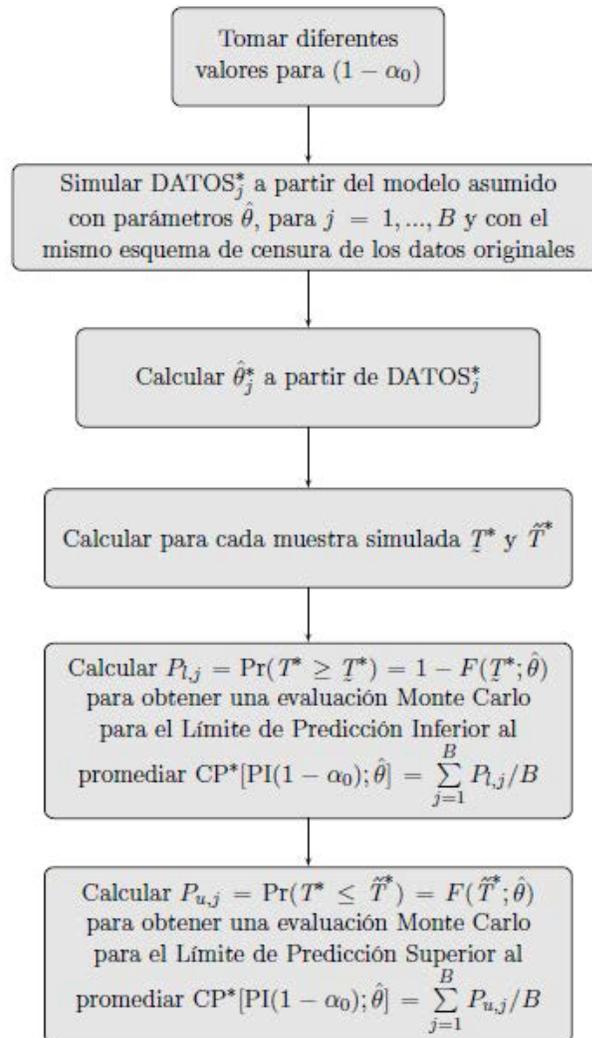


Figura 5: Algoritmo para la calibración de los límites de predicción inferior y superior promediando las probabilidades de cobertura condicional (Método 3)

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.

## 4 EJEMPLO ILUSTRATIVO CON DATOS SIMULADOS

El problema con el componente A que fue descrito anteriormente 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 artículo se obtendrán 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.

### 4.1 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 una distribución lognormal discretizada con media  $\mu = 1.6$  y desviación estándar  $\sigma = 0.5$  (ver Figura 1). En la Tabla 1 se presentan los valores para  $\mathbf{R}$  y  $\boldsymbol{\pi}$  obtenidos a partir de la simulación anterior.

Tabla 1: Tasas de uso simuladas del componente A en los hogares y sus probabilidades correspondientes

$R_i$	$\pi_i$	$R_i$	$\pi_i$
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  $\eta_C$  y  $\sigma_C$ .

Bajo la distribución de los ciclos hasta la falla dada en (2), se generó en R (2011) una muestra de tamaño  $m = 200$  con valores  $\eta_C = 200$  y  $\sigma_C = 1$ ; ver Tabla 2.

Tabla 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

Para obtener la muestra correspondiente al tiempo de vida del producto en campo (semanas de servicio), de acuerdo al modelo en (7), 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 y Tabla 1). Éste es un caso típico de generación de variables aleatorias discretas, por lo tanto, se utilizó el paquete estadístico R (2011) 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, para esto se tomaron nuevamente los valores de  $\eta_C = 200$  y  $\sigma_C = 1$ . En la Tabla 3 se pueden ver los valores obtenidos.

Tabla 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

## 4.2 Estimación por máxima verosimilitud para el modelo de tasa de uso con datos de vida censurados

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. El escenario que se va a considerar para el modelo de tasa de uso es aquel que se presenta cuando durante la ejecución de una ALT, el tiempo para observar las fallas es muy largo, por lo que resulta necesario limitarlo dando lugar a observaciones censuradas. Ésta es una censura a derecha no aleatoria debido a que la prueba se realiza hasta un número de ciclos predeterminado, por lo tanto la cantidad de unidades que no hayan fallado hasta este número de ciclos 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 tomó un valor para el número de ciclos a partir de los cuales quedarían censuradas las unidades en prueba, este valor corresponde al percentil 70 de la distribución de los  $m = 200$  ciclos hasta la falla lognormales generados en un principio. Dicho valor corresponde a 381 ciclos.

La expresión para la función de densidad de probabilidad (pdf) para los ciclos hasta la falla determinada a partir de una prueba de laboratorio (ALT) es lognormal y está dada por

$$\begin{aligned} f_{ALT}(c; \eta_C, \sigma_C) &= \frac{\partial F_C(c; \eta_C, \sigma_C)}{\partial c} \\ &= \frac{1}{c \sigma_C} \phi_{nor} \left[ \frac{\log(c) - \log(\eta_C)}{\sigma_C} \right], \end{aligned} \quad (15)$$

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

$$\begin{aligned} L_{ALT}(c; \eta_C, \sigma_C) &= \prod_{j=1}^m \left\{ \left( \frac{1}{c_j \sigma_C} \phi_{nor} \left[ \frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{\tau_j} \right\} \times \\ &\quad \left\{ \left( 1 - \Phi_{nor} \left[ \frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{1-\tau_j} \right\}, \end{aligned} \quad (16)$$

donde  $\tau_j$  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 en la ALT (ciclos hasta el fin de la prueba sin observar falla).

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

$$\begin{aligned} \mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C) = & \sum_{j=1}^m \tau_j \left\{ -\log(c_j) - \log(\sigma_C) + \log \left( \phi_{\text{nor}} \left[ \frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right) \right\} \\ & + (1 - \tau_j) \left\{ \log \left( 1 - \Phi_{\text{nor}} \left[ \frac{\log(c_j) - \log(\eta_C)}{\sigma_C} \right] \right) \right\}. \end{aligned} \quad (17)$$

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 las unidades en campo también 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 tomó un valor para determinar el tiempo de censura  $t_c$  del producto en manos del cliente, el percentil 70 de la distribución de los  $n = 100$  tiempos de vida en campo generados a partir de la mezcla de lognormales.

Se utilizó la función **norMix** con la cual se creó la mezcla univariada de normales sobre los valores  $\log(\eta_C/R_i)$ , con parámetro de dispersión  $\sigma_C$  y vector de probabilidades  $\boldsymbol{\pi}$  (ver ecuación (7)). Con la función **qnorMix** se calcularon los percentiles de la mezcla y posteriormente éstos fueron llevados a la escala original exponenciando los valores obtenidos por defecto de la función, para aproximar los percentiles correspondientes en la escala original de los datos.

El percentil encontrado, que corresponde al tiempo de censura,  $t_c$ , en unidades de las semanas y años de servicio del producto en campo, se presenta en la Tabla 4.

Tabla 4: Tiempos de censura en campo

Percentil	Tiempo de censura en semanas ( $t_c$ )	Tiempo de censura en años ( $t_c$ )
70	73	1.5

La función de densidad de probabilidad (pdf) para la mezcla finita de lognormales 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

$$\begin{aligned} f_{\text{Campo}}(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) &= \frac{\partial F(t; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R})}{\partial t} \\ &= \sum_{i=1}^k \frac{\pi_i}{t \sigma_C} \phi_{\text{nor}} \left[ \frac{\log(R_i \times t) - \log(\eta_C)}{\sigma_C} \right], \end{aligned} \quad (18)$$

donde cada valor de  $R_i$  representa la tasa de uso promedio del producto en manos del cliente en unidades de ciclos por semana y cada  $\pi$  es su probabilidad correspondiente (Meeker et al. 2009).

Los datos completos categorizados pueden ser expresados como

$$\mathbf{t}_{\text{comp}} = (\mathbf{t}^T, \mathbf{z}^T)^T, \quad (19)$$

donde

$$\mathbf{z} = (\mathbf{z}_1^T, \mathbf{z}_n^T)^T, \quad (20)$$

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 (Titterigton et al. (1985)).

Por lo tanto, 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

$$L_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \prod_{j=1}^n \left\{ \sum_{i=1}^k \pi_i^{z_{ij}} \left( \phi_{\text{nor}} \left[ \frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}} \right\}^{\delta_j} \\ \times \left\{ 1 - \sum_{i=1}^k \pi_i^{z_{ij}} \left( \Phi_{\text{nor}} \left[ \frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{z_{ij}} \right\}^{1-\delta_j}, \quad (21)$$

donde  $\delta_j$  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

$$\mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \sum_{j=1}^n \delta_j \log \left\{ \sum_{i=1}^k \pi_i^{z_{ij}} \left( \phi_{\text{nor}} \left[ \frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \frac{1}{t_j \sigma_C} \right)^{z_{ij}} \right\} \\ + (1 - \delta_j) \log \left\{ 1 - \sum_{i=1}^k \pi_i^{z_{ij}} \left( \Phi_{\text{nor}} \left[ \frac{\log(R_i \times t_j) - \log(\eta_C)}{\sigma_C} \right] \right)^{z_{ij}} \right\}. \quad (22)$$

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

$$\mathcal{L}(t_{\text{comp}}, c; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) = \mathcal{L}_{\text{Campo}}(t_{\text{comp}}; \eta_C, \sigma_C, \boldsymbol{\pi}, \mathbf{R}) + \mathcal{L}_{\text{ALT}}(c; \eta_C, \sigma_C) \quad (23)$$

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 (23). Para esto se puede recurrir a métodos como el de Newton-Rapson, Nelder-Mead o al método conocido como Evolución Diferencial (Beyer y Schwefel, 2000), este último implementado en el paquete estadístico R (2011) en la librería **DEoptim**.

En la Tabla 5 se presenta la estimación para  $\eta_C$  y  $\sigma_C$ , 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**.

Tabla 5: Parámetros estimados vía Evolución Diferencial para los datos de vida del componente A censurados.

Tiempo de censura a derecha	Parámetro	Estimación
Percentil 70	$\hat{\eta}_C$	190.0611496
	$\hat{\sigma}_C$	0.6683192

### 4.3 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 censurados, el paso a seguir es el cálculo de los intervalos de predicción bilaterales y los límites de predicción unilaterales naive.

Un intervalo de predicción bilateral naive para  $T$  al  $(1 - \alpha)100\%$  está dado por

$$PI(1 - \alpha) = [T, \tilde{T}] = [\hat{t}_{\alpha/2}, \hat{t}_{1-\alpha/2}], \quad (24)$$

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

$$PI(1 - \alpha_l) = [T, \infty) = [\hat{t}_\alpha, +\infty). \quad (25)$$

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

$$PI(1 - \alpha_u) = (0, \tilde{T}] = (0, \hat{t}_{1-\alpha}]. \quad (26)$$

Para calcular estos intervalos y límites de predicción para el modelo de tasa de uso propuesto por Meeker et al. (2009) dado en (6), 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 **norMix** y **qnorMix** de la librería **nor1mix** del paquete estadístico R (2011) para aproximar estos valores.

A continuación se presentan los intervalos de predicción bilaterales  $PI(1 - \alpha)$  naive y los límites de predicción unilaterales inferiores  $PI(1 - \alpha_l)$  y superiores  $PI(1 - \alpha_u)$  naive con probabilidades de cobertura correspondientes  $CP[PI(1 - \alpha); \hat{\theta}]$ ,  $CP[PI(1 - \alpha_l); \hat{\theta}]$  y  $CP[PI(1 - \alpha_u); \hat{\theta}]$  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 censura:

Tabla 6: Intervalos de predicción bilaterales y límites de predicción unilaterales naive

	Bilaterales	Unilaterales inferiores	Unilaterales superiores
CP[PI(1 - $\alpha$ ); $\hat{\theta}$ ]	PI(1 - $\alpha$ )	PI(1 - $\alpha_l$ )	PI(1 - $\alpha_u$ )
0.90	[9.828942, 153.556945]	[13.26696, + $\infty$ )	(0, 112.84895]
0.95	[7.585515, 200.990628]	[9.828942, + $\infty$ )	(0, 153.556945]

Como puede verse en las Tablas anteriores, 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  $\Pr(0 < T \leq \tilde{T}) = 1 - \alpha/2$ , entonces  $\Pr(\underline{T} \leq T \leq \tilde{T}) = 1 - \alpha$ . 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 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  $\eta_C$  y  $\sigma_C$ , 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.

#### 4.3.1 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 el **Método 1**, descrito en el algoritmo de la Figura 2 y la Figura 3 y el **Método 3**, descrito en el algoritmo de la Figura 5, ambos con un tamaño de muestra en la simulación de  $B = 5.000$ . El **Método 2**, que se describe en la Figura 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 (6) evaluada en los parámetros estimados de acuerdo al esquema de censura establecido, así

$$F(t; \hat{\eta}_C, \hat{\sigma}_C, \boldsymbol{\pi}, \mathbf{R}) = \Pr(T \leq t) = \sum_{i=1}^k \pi_i \Phi_{nor} \left[ \frac{\log(t) - \log(\hat{\eta}_C/R_i)}{\hat{\sigma}_C} \right]. \quad (27)$$

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 (2011) (ver Durbán, 2009). Además, la

probabilidad de cobertura que se tomó para la construcción de los intervalos y límites de predicción calibrados fue de 0.95.

Finalmente, para calcular los límites de predicción unilaterales sólo se necesitó encontrar el valor  $1 - \alpha_c$  apropiado. Debido a que  $CP^*[PI(1 - \alpha_c); \hat{\theta}]$  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.

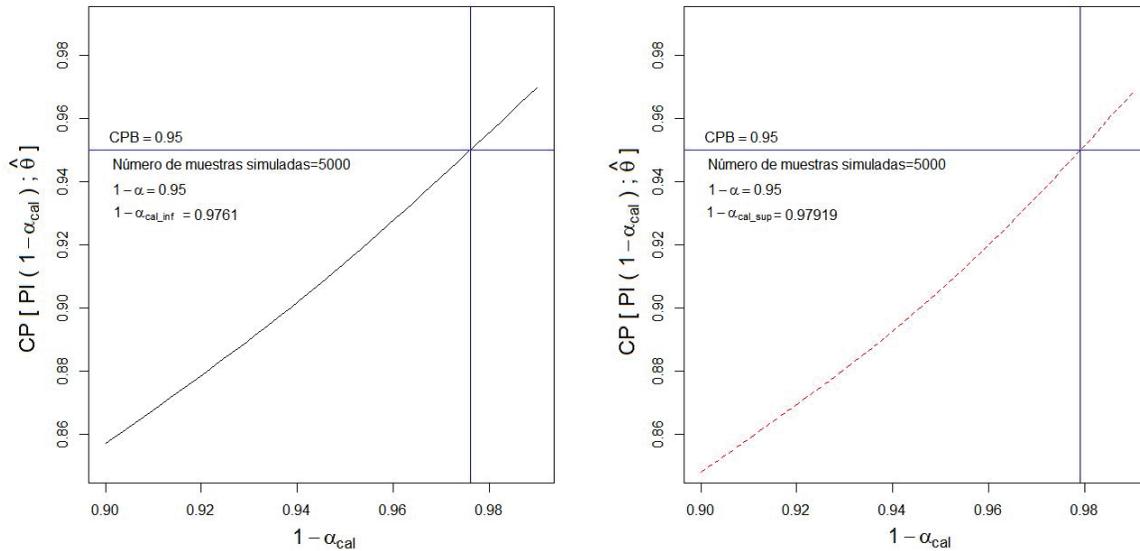


Figura 6: a. Curva de calibración del límite inferior por el Método 1. b. Curva de calibración del límite superior por el Método 1.

Usando los puntos de calibración con el Método 1 en la Figura 6, 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.

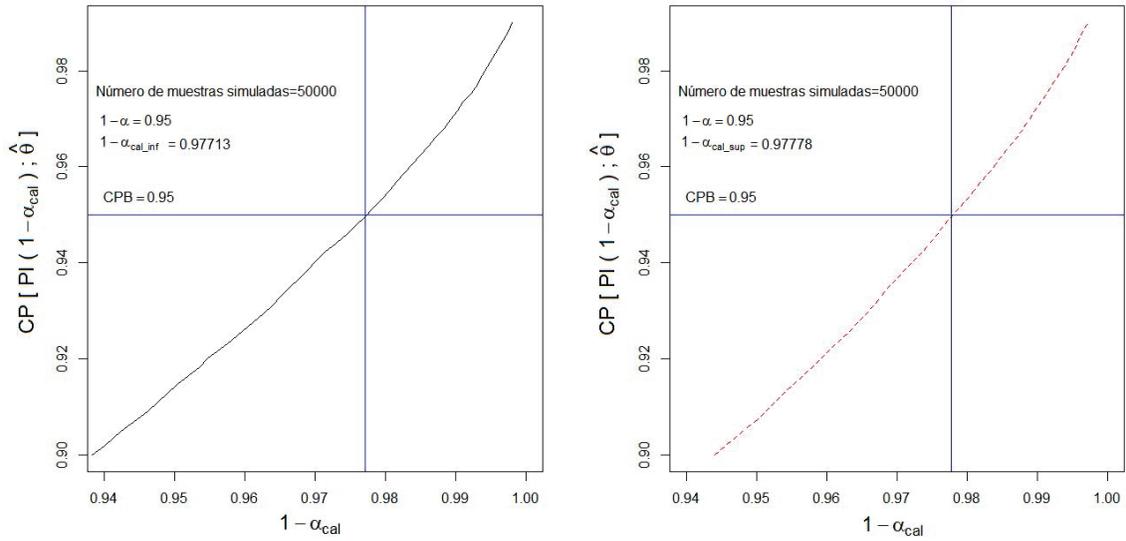


Figura 7: a. Curva de calibración del límite inferior por el Método 2. b. Curva de calibración del límite superior por el Método 2.

Usando los puntos de calibración con el Método 2 en la Figura 7, 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.

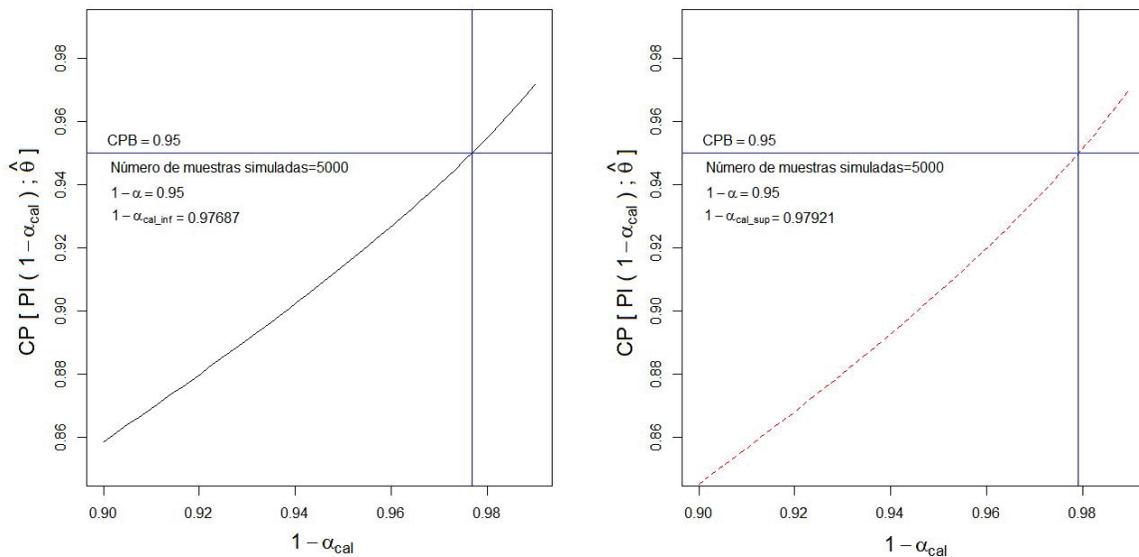


Figura 8: a. Curva de calibración del límite inferior por el Método 3. b. Curva de calibración del límite superior por el Método 3.

Usando los puntos de calibración con el Método 3 en la Figura 8, 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:

Tabla 7: Límites de predicción unilaterales inferiores calibrados

<b>CP[PI(1 - <math>\alpha_{cl}</math>)] = 0.90</b>			
	Límite inferior calibrado	Límite inferior calibrado	Límite inferior calibrado
Límite inferior naive	Método 1	Método 2	Método 3
[13.26696, +∞)	[10.691176, +∞)	[10.713769, +∞)	[10.729732, +∞)
<b>CP[PI(1 - <math>\alpha_{cl}</math>)] = 0.95</b>			
[9.828942, +∞)	[7.466977, +∞)	[7.354432, +∞)	[7.382321, +∞)

Tabla 8: Límites de predicción unilaterales superiores calibrados

<b>CP[PI(1 - <math>\alpha_{cu}</math>)] = 0.90</b>			
	Límite superior calibrado	Límite superior calibrado	Límite superior calibrado
Límite superior naive	Método 1	Método 2	Método 3
(0, 112.84895]	(0, 148.4274]	(0, 146.5511]	(0, 148.2265]
<b>CP[PI(1 - <math>\alpha_{cu}</math>)] = 0.95</b>			
(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%, los límites de predicción unilaterales naive al 95% pueden ser combinados, como se muestra en la Tabla 9.

Tabla 9: Intervalos de predicción bilaterales calibrados

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

De acuerdo con estos resultados, 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.

## 5 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) y con base en los escenarios considerados, 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 (Meeker et al. (2009)).
- Los intervalos de predicción estimados por el procedimiento naive son 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  $\theta$ . 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.
- 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 ( $B = 50000$ ).

## Referencias

- Atwood, C. L. (1984), Approximate tolerance intervals based on maximum likelihood estimates. *Journal of the American Statistical Association*, **79**, 459-465.
- Beyer, H. G; Schwefel, H. P. (2002), Evolution strategies -A comprehensive introduction. *Natural Computing*, **1**, 3-52.
- Cox, D. R. (1975), Prediction intervals and empirical Bayes confidence intervals. *Perspectives in Probability and Statistics*, 47-55.
- Durbán, M. (2009), An introduction to smoothing with penalties: P-splines. *Boletín de Estadística e Investigación Operativa*, **25**, 195-205.
- Escobar, L. A.; Meeker, W. (1999), Statistical prediction based on censored life data. *Technometrics*, **41**, 113-124.
- Mee, R.; Kushary, D. (1994), Prediction limits for the weibull distribution utilizing simulation. *Computational Statistics and Data Analysis*, **17**, 327-336.
- Meeker, W.; Escobar, L. A. (1998), *Statistical methods for reliability data*. New York: Wiley.
- Meeker, W.; Escobar, L. A.; Hong, Y. (2009), Using accelerated life tests results to predict product field reliability. *Technometrics*, **51**, 146-161.
- Morgan, B. J. T. (1984), *Elements of simulation*. Chapman and Hall.
- R Development Core Team (2010), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- Titterigton, D. M.; Smith, A. F. M.; Makov, U. E. (1985), *Statistical analysis of finite mixture distributions*. John Wiley and Sons, Inc.