

ESTIMACIÓN BAYESIANA PARA LA FUNCIÓN DE CONFIABILIDAD

(BAYESIAN ESTIMATION FOR RELIABILITY FUNCTION)

Recibido: 23/10/2016

Aprobado: 27/12/2016

Javier Ramírez Montoya

Departamento de Matemáticas y Estadística, Grupo de Investigación en Estadística
Universidad de Córdoba, Montería, Colombia
jaramirez@correo.unicordoba.edu.co

José Martínez Salazar

Facultad de Ciencias Básicas e Ingenierías, Grupo de Investigación TESEEO
Universidad del Sinú, Montería, Colombia
jmartinez@unisinu.edu.co

RESUMEN

En el área de la industria, en particular en la ingeniería, la confiabilidad de productos o sistemas es fundamental para la buena eficiencia de los mismos. En este artículo, se estima la función de confiabilidad utilizando estadística bayesiana, involucrando *a priori* no informativa. Además, se tiene la estimación paramétrica vía máxima verosimilitud para el caso exponencial. Para ello, se generan muestras artificiales mediante simulación de Monte Carlo, considerando diferentes tamaños de muestra y porcentajes de censura. La estimación se implementa con diferentes métodos y se toma como referencia la estimación bayesiana dado un modelo de confiabilidad teórico, utilizando como medida de calidad de la estimación el error medio cuadrático. Los resultados obtenidos señalan que el estimador no paramétrico de Nelson-Aalen fue la estrategia más eficiente, independientemente del tamaño de muestra y porcentaje de censura tomados en el estudio, mientras que la estimación bayesiana utilizando *a priori* no informativa para los tiempos observados y censurados distribuidos de manera exponencial no presentó los mejores resultados, siendo la censura un fenómeno altamente determinante en la estimación de la función de confiabilidad y el tamaño adecuado de la muestra.

Palabras claves: Estadística Bayesiana, Función de confiabilidad, Estimador no paramétrico de Nelson-Aalen.

ABSTRACT

In the area of industry, particularly in engineering, the reliability of products or systems is fundamental for the good efficiency of them. In this article, the reliability function is estimated using Bayesian statistics, involving a non-informative *a priori*. In addition, we find the parametric estimation via maximum likelihood for the exponential case. To do this, artificial samples are generated by Monte Carlo simulation, considering different sample sizes and censorship percentages. The estimation is implemented with different methods and the

Bayesian estimate is taken as a reference given a theoretical reliability model, using the mean squared error as the measure of quality of the estimate. The results obtained indicate that the non-parametric Nelson-Aalen estimator was the most efficient strategy, regardless of the sample size and censorship percentage in the study, while the Bayesian estimate using a priori non-informative for the observed and censored times distributed exponentially it did not present the best results, being censorship a highly determinant phenomenon in the estimation of the reliability function and the adequate size of the sample.

Keywords: *Bayesian statistics, Reliability function, non-parametric Nelson-Aalen estimator.*

INTRODUCCIÓN

En Ingeniería, el tiempo de falla o de ocurrencia de un fenómeno de interés puede ser modelado mediante análisis de confiabilidad, siendo la estadística bayesiana ampliamente utilizada en particular con a priori no informativas, constituyéndose en una alternativa interesante cuando se conoce la distribución de dichos tiempos así como de la censura. Además, el algoritmo algebraico de la mencionada estadística es relativamente sencillo de implementar [1].

En la literatura, se reportan diversos estudios y aplicaciones que consideran la estadística bayesiana utilizando la distribución exponencial [2-6] y en los cuales se observa la importancia de evaluar cómo los tamaños de muestra y porcentajes de censura afectan la estimación. En este artículo, se analiza el impacto de dichos parámetros en la referida estimación para los casos de tamaños de muestra muy pequeños, que son habituales en estudios de ingeniería [7]. Para ello, se recurre a la simulación Monte Carlo para comparar el estimador bayesiano en el caso de a priori no informativa, con el método paramétrico de máxima verosimilitud, así como con métodos no paramétricos, para de esa manera ampliar los resultados reportados por Ramírez y Martínez [6].

En la siguiente sección se presenta la función de interés a estimar y los métodos utilizados para llevar a cabo dicha estimación. Luego, son analizados los escenarios de la simulación, para finalmente presentar los resultados obtenidos.

ESTIMACIÓN DE LA FUNCIÓN DE CONFIABILIDAD

Se considera que la función de confiabilidad, S , corresponde a la probabilidad de que un sistema o componente ejecute sus funciones bajo condiciones de operación, por un periodo específico de tiempo, es decir [8]:

$$S(t) = P(T > t) \quad (1)$$

Tal como se aprecia en (1), la función de confiabilidad es el complemento de una función de distribución acumulativa, por tanto, es decreciente con el tiempo. El tiempo total es conformado por n tiempos, de los cuales $k < n$ son censurados.



Estimación Bayesiana

Con base en el análisis bayesiano objetivo, teniendo en cuenta distribuciones no informativas, el teorema de Bayes, para las variables aleatorias X y θ , cuyas funciones de densidades de probabilidad son $f(\theta|x)$ y $\xi(\theta)$, establece que:

$$\xi(\theta) = \frac{f(x|\theta) \xi(\theta)}{\int_{\theta} f(\theta|x) \xi(\theta) d\theta} \quad (2)$$

En el contexto bayesiano, X corresponde a los datos (escalar o vector o matriz), θ es el parámetro desconocido, $f(x_1, \dots, x_n|\theta)$ corresponde a la verosimilitud de los datos dado el parámetro θ y $\xi(\theta)$ es la distribución a priori de θ .

Por tanto, por el teorema de Bayes, la distribución a posteriori es dada por [9]:

$$\xi(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta) \xi(\theta)}{\int_{\theta} f(\theta|x_1, \dots, x_n) \xi(\theta) d\theta} \quad (3)$$

Así, la regla de Bayes se escribe como:

$$\xi(\theta|x_1, \dots, x_n) \propto f(x_1, \dots, x_n|\theta) \xi(\theta) \quad (4)$$

Luego, para satisfacer la propiedad local de uniformidad para distribuciones a priori no informativas, se considera la regla general propuesta por Jeffreys [10]:

$$I(\theta) = -E \left[\frac{\partial^2 \log(f(x|\theta))}{\partial \theta^2} \right] \quad (5)$$

Siendo la a priori definida como:

$$\xi(\theta) \propto |I(\theta)|^2 \quad (6)$$

Para el caso de distribución exponencial para los tiempos, es decir, función de densidad de probabilidad $f(t) = \lambda e^{-\lambda t}$, se tiene, de (5):

$$I(\lambda) = -E \left[\frac{\partial^2 \log(f(x|\lambda))}{\partial \lambda^2} \right] \quad (7)$$

Donde:

$$\log(f(t|\lambda)) = \log(\lambda) - \lambda t \quad (8)$$

Por lo que, de (8) se obtiene:



$$\frac{\partial^2 \log(f(x|\lambda))}{\partial \lambda^2} = -\frac{1}{\lambda^2} t \quad (9)$$

Por tanto:

$$I(\lambda) = -E \left[-\frac{1}{\lambda^2} \right] = \frac{1}{\lambda^2} \quad (10)$$

Por otra parte, considerando r tiempos observados y $(n - r)$ tiempos censurados, entonces:

$$L(x, \lambda) = \prod_{i=1}^r \lambda e^{-\lambda t_i} \prod_{j=1}^{n-r} \lambda e^{-\lambda t_j} = \lambda^r e^{-\lambda \sum_{i=1}^r t_i} e^{-\lambda \sum_{j=1}^{n-r} t_j} = \lambda^r e^{-n\lambda \bar{t}} \quad (11)$$

Obteniéndose la distribución a posteriori dada por:

$$\xi(\lambda|t) = I(\lambda)L(x, \lambda) = \lambda^{-2} \lambda^r e^{-n\lambda \bar{t}} = \lambda^{r-2} e^{-n\lambda \bar{t}} \quad (12)$$

Para una distribución a posteriori con los tiempos distribuidos exponencialmente, se tiene:

$$\log(\xi(\lambda|t)) = (r - 2)\log(\lambda) - n\lambda \bar{t} \quad (12)$$

De (12):

$$\frac{\partial \log(\xi(\lambda|t))}{\partial \lambda} = \frac{(r-2)}{\lambda} - n\bar{t} \quad (13)$$

Igualando a cero la ecuación (13) se obtiene que el estimador bayesiano es dado por:

$$\hat{\lambda}_{BAY} = \frac{(r-2)}{\sum_{i=1}^n t_i} \quad (14)$$

Cuya varianza es:

$$\text{Var}(\hat{\lambda}) = \frac{(r-2)^2}{\frac{n}{\lambda^2}} = \frac{\lambda^2(r-2)^2}{n} \quad (15)$$

Ahora bien, considerando que la función de confiabilidad es dada, de forma general, por la ecuación (1), y asumiendo un tiempo de interés t_0 , se considerarán los siguientes estimadores para su comparación [11, 12]:



Bayesiano

$$\hat{S}(t_0) = e^{-\hat{\lambda}_{MV}t_0} = e^{-\frac{(r-2)}{\sum_{i=1}^n t_i}t_0} \quad (16)$$

Máximo verosímil

$$\hat{S}(t_0) = e^{-\hat{\lambda}_{BAY}t_0} = e^{-\frac{r}{\sum_{i=1}^n t_i}t_0} \quad (17)$$

No paramétrico Kaplan-Meier

$$\hat{S}(t_0) = \prod_{j:t_j < t_0} \frac{(n_j - d_j)}{n_j} \quad (18)$$

No paramétrico Nelson-Aalen

$$\hat{S}(t_0) = e^{-H(t_0)} \quad (19)$$

Para comparar la eficiencia de cada uno de los mencionados, se utiliza como métrica el error medio cuadrático dado por [12]:

$$[\hat{S}(t_0) - S(t_0)]^2 + \text{Var}(\hat{S}(t_0)) \quad (20)$$

SIMULACIÓN Y RESULTADOS

Con el propósito de estimar la función de confiabilidad mediante estimación bayesiana y comparar su eficiencia con la arrojada por el estimador paramétrico tradicional de máxima verosimilitud y dos de los estimadores no paramétricos, se generan muestras con base en tiempos de falla/censura distribuidos de manera exponencial. La estimación se lleva a cabo mediante simulación de Monte Carlo, calculando los promedios de los errores medios cuadráticos de cada estimador teniendo en cuenta tamaños de muestra $n = (10, 15, 20)$ y porcentajes de censura (0%, 15%, 30%), generando 2000 muestras.

El tiempo de interés, t_0 , considerado es de 0.2 segundos, mientras que la función de confiabilidad teórica asumida es $S(t) = e^{-\lambda t}$, con $\lambda = (0.5, 1, 2)$. Por su parte, para r tiempos observados se considerarán los valores 1 y 2, recordando que el tiempo generado es dado por $e^{(r, \lambda)}$.

Resultados

A modo de ejemplo, en las Tablas 1 y 2, se resumen los valores del error medio cuadrático arrojado por cada uno de los estimadores considerados, para $n = 10$ y tiempos generados $e^{(1, 0.5)}$ y $e^{(2, 0.5)}$, respectivamente. En las mencionadas tablas, Bay, KM, NA y MV



se refieren a los estimadores bayesiano, Kaplan-Meier, Nelson-Aalen y máxima verosimilitud, en ese orden.

Tabla 1. Promedios del error medio cuadrático, $n = 10$, $e^{(1,0.5)}$

Porcentaje de censura	Método estimador			
	Bay	NA	KM	MV
0	0.0953	0.0362	0.0412	0.0729
15	0.0729	0.0290	0.0326	0.0543
30	0.0623	0.0270	0.0302	0.0462

Tabla 2. Promedios del error medio cuadrático, $n = 10$, $e^{(2,0.5)}$

Porcentaje de censura	Método estimador			
	Bay	NA	KM	MV
0	0.2273	0.0449	0.0513	0.1487
15	0.1487	0.0357	0.0392	0.1108
30	0.1315	0.0342	0.0364	0.1024

De las Tablas 1 y 2, se observa que el estimador con mayor eficiencia es el de Nelson-Aalen (NA), presentando el menor error medio cuadrático para todos los tres porcentajes de censura considerados, mientras que el estimador bayesiano (Bay) es el que arrojó los valores de error medio cuadrático más elevados.

Los resultados reportados en las Tablas 1 y 2 son ilustrados en las Figuras 1a y 1b, respectivamente, en las cuales se muestra el desempeño de cada estimador en términos del error medio cuadrático obtenido en función del porcentaje de censura.

En las Figuras 1a y 1b, se aprecia que a medida que aumenta el porcentaje de censura también aumenta el error medio cuadrático, lo cual es debido a la influencia negativa que tiene la censura sobre la estimación. Adicionalmente, se observa un agrupamiento de los estimadores no paramétricos, con mejor desempeño en la estimación que el agrupamiento de los estimadores bayesiano y máxima verosimilitud, los cuales arrojan el desempeño más bajo. Esto resultados eran de esperarse teniendo en cuenta el trabajo de [10], en el cual se consideró el mismo tipo de censura a la derecha y la distribución de Pareto.

En las Figuras 1a y 1b, también se puede observar que a medida que aumenta la cantidad de tiempos observados r , que conlleva a una disminución de los tiempos censurados ($n - r$), la diferencia entre los valores del error medio cuadrático arrojados por los estimadores no paramétricos, es decir, Nelson-Aalen (NA) y Kaplan-Meier (KM) disminuye.

Otro resultado interesante de ser analizado es el que se obtiene aumentando el tamaño n , de la muestra. En ese sentido, en las Tablas 3 y 4 se resumen los valores del error medio cuadrático arrojados por los estimadores cuando n es igual a 15, considerando los mismos tiempos generados del caso anterior, es decir, $e^{(1,0.5)}$ y $e^{(2,0.5)}$, respectivamente, así como los mismos porcentajes de censura.



Figura 1. Promedios del error medio cuadrático, $n = 10$: a) Tiempo generado $e^{(1,0.5)}$; b) Tiempo generado $e^{(2,0.5)}$

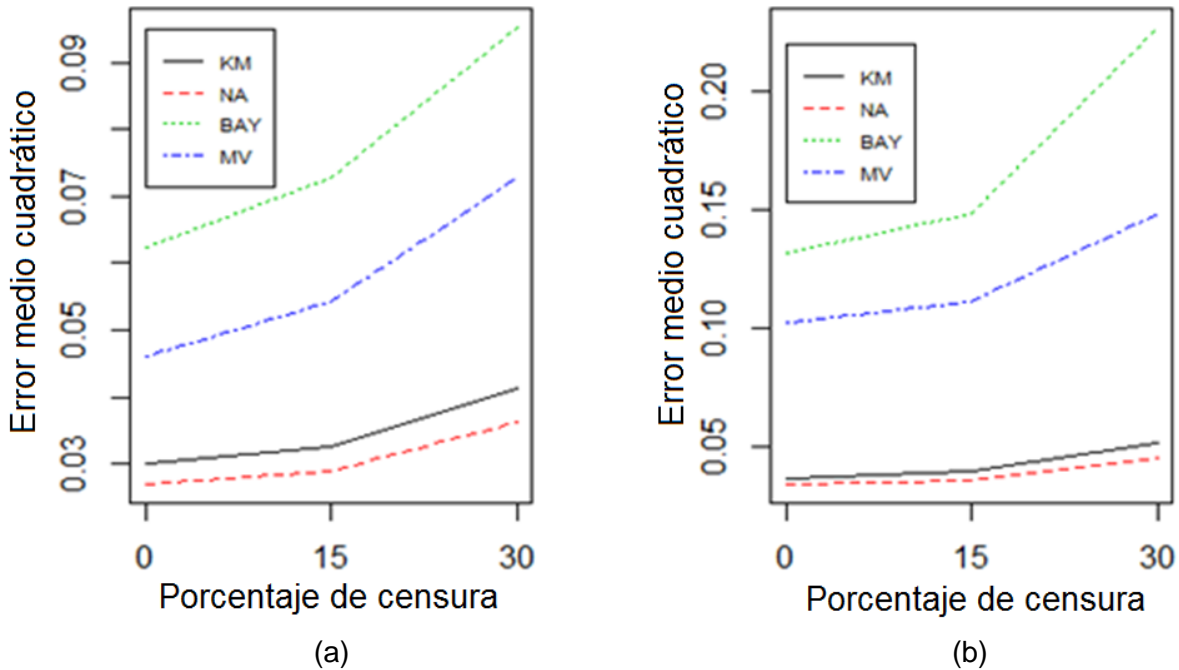


Tabla 3. Promedios del error medio cuadrático, $n = 15$, $e^{(1,0.5)}$

Porcentaje de censura	Método estimador			
	Bay	NA	KM	MV
0	0.0865	0.0226	0.0247	0.0729
15	0.0729	0.0197	0.0214	0.0606
30	0.0536	0.0174	0.0183	0.0441

Tabla 4. Promedios del error medio cuadrático, $n = 15$, $e^{(2,0.5)}$

Porcentaje de censura	Método estimador			
	Bay	NA	KM	MV
0	0.1956	0.0299	0.0328	0.1487
15	0.1487	0.0260	0.0277	0.1206
30	0.1193	0.0256	0.0259	0.1026

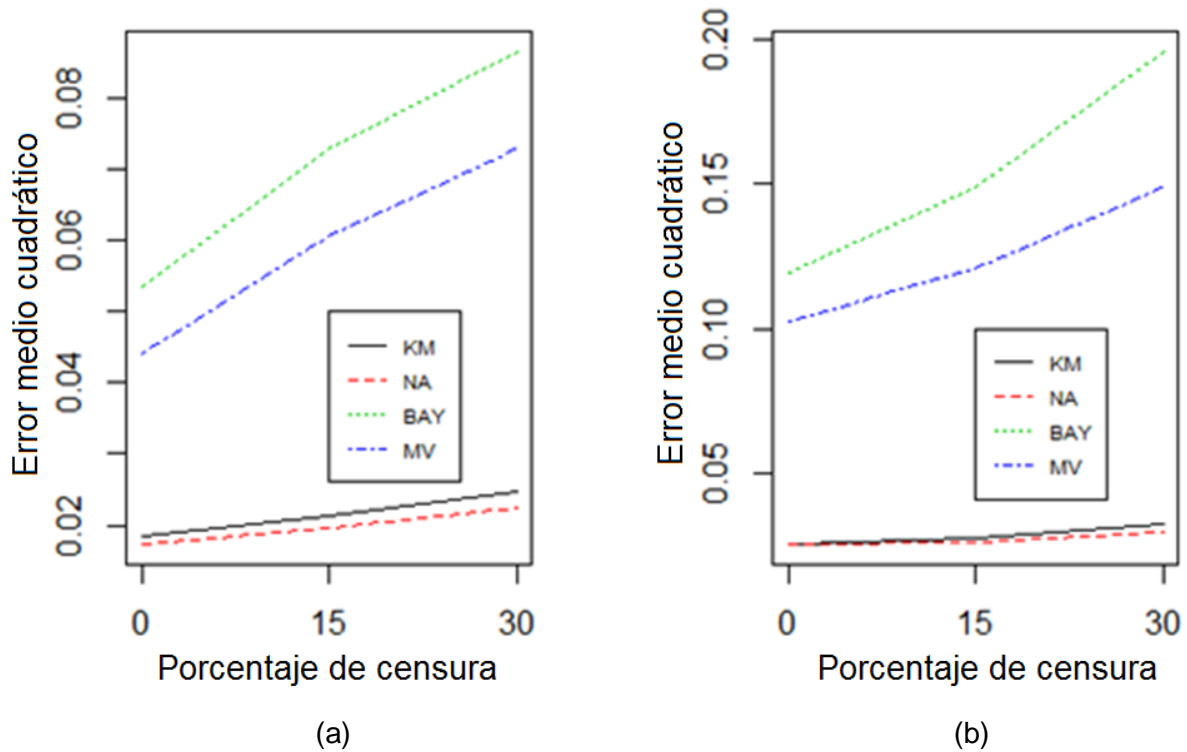
En las Tablas 3 y 4, nuevamente se aprecia el mejor desempeño, para todos los tres porcentajes de censura considerados, del estimador Nelson-Aalen (NA). Asimismo, se observa que de nuevo la estimación bayesiana es la que presenta el desempeño más bajo, es decir, arroja los mayores valores de error medio cuadrático.

La ilustración gráfica de los resultados que aparecen en las Tablas 3 y 4, se puede observar en las Figuras 2a y 2b, en las cuales se aprecia que la separación entre los dos grupos de tendencia, es decir, el de menores valores del error medio cuadrático



(estimadores no paramétricos) y el de mayores valores de dicho error (estimadores bayesiano y máxima verosimilitud), aumenta.

Figura 2. Promedios del error medio cuadrático, $n = 15$: a) Tiempo generado $e^{(1,0.5)}$; b) Tiempo generado $e^{(2,0.5)}$



Finalmente, la simulación se repite, pero aumentando el valor de la muestra a 20. Los resultados obtenidos aparecen en las Tablas 5 y 6 y se ilustran en las Figuras 3a y 3c.

Tabla 5. Promedios del error medio cuadrático, $n = 20$, $e^{(1,0.5)}$

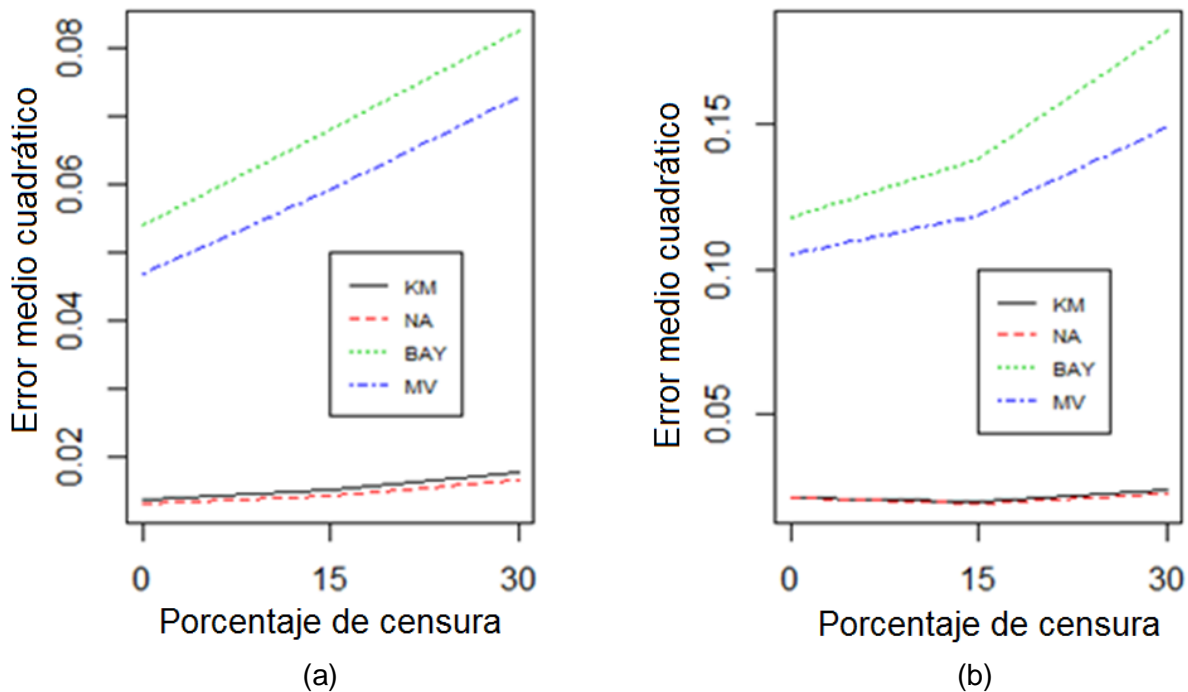
Porcentaje de censura	Método estimador			
	Bay	NA	KM	MV
0	0.0827	0.0166	0.0178	0.0729
15	0.0680	0.0143	0.0151	0.0593
30	0.0541	0.0132	0.0137	0.0469

Tabla 6. Promedios del error medio cuadrático, $n = 20$, $e^{(2,0.5)}$

Porcentaje de censura	Método estimador			
	Bay	NA	KM	MV
0	0.1824	0.0224	0.0240	0.1487
15	0.1381	0.0193	0.0201	0.1188
30	0.1182	0.0211	0.0210	0.1051

Se observa, una vez más, que el estimador que arrojó los menores valores de error medio cuadrático es el de Nelson-Aalen, mientras que los mayores valores del mencionado error son los producidos por el estimador bayesiano. Asimismo, se puede apreciar que la separación entre los dos (2) grupos de estimadores arriba mencionados, es decir, el grupo conformado por los estimadores bayesiano y máxima verosimilitud y el grupo integrado por los estimadores no paramétricos, es mayor en comparación con n igual a 15.

Figura 2. Promedios del error medio cuadrático, $n = 20$: a) Tiempo generado $e^{(1,0.5)}$; b) Tiempo generado $e^{(2,0.5)}$



CONCLUSIONES

Los métodos de estimación bayesiana con *a priori* no informativa para modelar tiempos de falla y censura distribuidos exponencialmente no presentan buenos resultados en la estimación de la función de confiabilidad con tamaños de muestra pequeños; sin embargo, otros autores actualmente muestran mejores resultados con distribuciones Rayleigh [11].

Los estimadores no paramétricos pueden ser más apropiados con tamaños de muestra pequeños y porcentajes de censura bajos. En particular, el estimador de Nelson-Aalen es recomendado en los escenarios tomados en este trabajo, coincidiendo con los resultados de [6].

REFERENCIAS BIBLIOGRÁFICAS

- [1] Bambeý C. (2014). Bayesian Parameter and Reliability Estimate of Weibull Failure Time Distribution. *Malasyan Mathematical Sciences Society*, Vol. 37, No.2 (Pp. 611-632).

- [2] Tsokos C. (1977). The Theory and Applications of Reliability with Emphasis on Bayesian and Nonparametric Methods. Volume 1. Academic Press. New York.
- [3] Pensky M. and Singh R. (1999). Empirical of Reliability Characteristics for an Exponential Family. The Canadian Journal of Statistics, Vol. 27, No. 1 (Pp. 127-136).
- [4] Solow A.R. (2016). On the Prior Distribution of Extinction Time. Biology Letters, Vol. 12, No. 6 (Pp. 1-4).
- [5] Thair M. and Aslam M. (2015). Bayesian Analysis of the 3-Component Mixture of Exponential Distribution Assuming the Non-Informative Priors. Revista Colombiana de Estadística, Vol. 38, No. 2 (Pp. 431-452).
- [6] Ramírez J y Martínez G. (2015). Comparación de Estimadores no Paramétricos frente a los Paramétricos para la Función de Confiabilidad. Revista Ingeniería, Investigación y Desarrollo, Vol. 15, No. 2 (Pp. 15-24).
- [7] Aguirre-Torres V. and Vara R. (2014). A Bayesian Analysis of Very Small Unreplicated Experiments. Quality and Reliability Engineering International (Special Issue on European Network for Business and Industrial Statistics 2012), Vol. 30, No. 3 (Pp. 413-426).
- [8] [5] Meeker W and Escobar, L. (1998). Statistical Models and Methods for Reliability Data. John Wiley & Sons.
- [9] Correa J. (2005). Estadística Bayesiana. Notas de Clase, Universidad Nacional de Colombia, Medellín.
- [10] Walter G., Aslett L.J.M. and Coolen F.P.A. (2016). Bayesian Nonparametric System Reliability using Sets of Priors. International Journal of Approximate Reasoning. Documento en línea. Disponible en <https://doi.org/10.1016/j.ijar.2016.08.005> (Consulta: septiembre 2016).
- [11] [8] Kaplan E. and Meier P. (1958). Estimation from Incomplete Observations. American Statistical Association, Vol. 53 (Pp. 457-481).
- [12] [9] Lawless J. (2003). Statistical Models and Methods for Lifetime Data. John Wiley & Sons.
- [13] [10] Rasheed H.A. and Al-Gazi N.A.A. (2014). Bayesian Estimation for Reliability Function of Pareto Type I Distribution under Generalized Square Error Loss Function. International journal of Engineering and Innovative Technology, Vol. 4, No. 6 (Pp. 33-40).

- [14] Ferreira J.T., Bekker A., Arashi M. (2016). Objective Bayesian Estimators for the Right Censored Rayleigh Distribution: Evaluating the Al-Bayyati Loss Function. REVSTAT-Statistical Journal, Vol. 14, No. 4 (Pp. 433-454).

BIOGRAFIA DE LOS AUTORES

Javier Ramírez Montoya

Es Licenciado en Estadística en la Universidad de Córdoba, Montería, Colombia. Magíster en Ciencias Estadísticas en la Universidad Nacional de Colombia, Medellín. Es Profesor Asociado en el Departamento de Matemáticas y Estadística de la Universidad de Córdoba, Montería, Colombia y Es miembro del Grupo de Investigación en Estadística de la misma universidad. Sus principales áreas de experticia son el Análisis de Supervivencia, Confiabilidad y Análisis de Datos Multivariados

José Martínez Salazar

Doctor en Estadística, en la Universidad Central de Venezuela, Caracas. Es Profesor Titular en la Universidad del Sinú, Montería, Colombia, adscrito al Departamento de Ciencias Básicas de la Facultad de Ciencias e Ingenierías. Es miembro del grupo de Investigación TESEEO, fungiendo como director de la línea de investigación Aplicación de las Matemáticas en el proceso de Análisis y Estructuración de Datos provenientes de un Fenómeno de Interés. Entre sus áreas de experticia están la Probabilidad, Modelos Lineales y Análisis de Datos Multivariados
