

CONFIABILIDAD DE DISPOSITIVO OBSERVABLE EN UN INSTANTE DEL TIEMPO: APLICACIÓN A COMPONENTES DE TRANSMISIÓN DE RADARES

OBSERVABLE RELIABILITY OF DEVICE IN A MOMENT OF TIME: APPLICATION TO DRIVE COMPONENTS RADAR

Juan Pablo Valencia Arango^{1,*}, Juan Carlos Correa Morales²

RESUMEN

Determinar la confiabilidad de dispositivos es una de las principales tareas de la ingeniería. En algunas situaciones experimentales sólo es posible observar el tiempo completo de falla de una unidad. Sin embargo, para chequear si está o no trabajando basta con un punto del tiempo. Bajo algunos modelos paramétricos se pueden desarrollar análisis estadísticos de estas situaciones encontrando los estimadores de máxima verosimilitud, sus distribuciones asintóticas y los intervalos de confianza. El presente trabajo tiene como objetivo presentar una propuesta metodológica para el cálculo del intervalo de confianza de un dispositivo en presencia de doble censura. Se presenta la metodología paramétrica bajo la distribución exponencial, mostrando de manera detallada el cálculo del intervalo de confianza para el tiempo medio de falla. Posteriormente se presentan simulaciones para el cálculo del nivel real y longitud media del intervalo de confianza, se encuentra que los niveles alcanzados en el caso de doble censura son muy similares a cuando se observan los tiempos completos. Los resultados son aceptables y recomendados en situaciones en las que es costosa la observación de la muestra a lo largo del tiempo.

Palabras Claves: Dato censurado, intervalo de confianza, distribución exponencial.

ABSTRACT

To determine the reliability of devices is one of the main tasks of engineering. In some experimental situations it is only possible to observe the full time of a drive failure. However, to check whether or not working just one point of time. Under some parametric models can develop statistical analyzes of these situations by finding the maximum likelihood estimators, their asymptotic distributions and confidence intervals. This paper aims to present a methodology for the calculation of the confidence interval of a device in the presence of double censorship. Parametric methodology under the exponential distribution is presented, showing in detail the calculation of the confidence interval for the mean time to failure. Later simulations to calculate the actual level and average length of the confidence interval are presented, it is that the levels reached in the case of double censorship are very similar to when full time was observed. The results are acceptable and recommended in situations where it is costly observing the sample over time.

Keywords: Censored data, confidence interval, exponential distribution.

¹Estudiante de Doctorado en Ingeniería. Universidad Nacional de Colombia, Sede Medellín. Docente de tiempo Completo Escuela de Ingeniería de Antioquia. Envigado, Colombia.

²PhD en Estadística, University of Kentucky. Docente asociado de la Escuela de Estadística. Universidad Nacional de Colombia, Sede Medellín.

*Autor para correspondencia: jpvale0@unal.edu.co

INTRODUCCIÓN

Los estudios de confiabilidad son de alta importancia en temas de ingeniería. Dentro de este tipo de estudio en muchas ocasiones se trabajan con datos censurados. Lagakos (1979) expresa que los datos censurados juegan un papel muy importante en diferentes aplicaciones, puesto que no siempre es posible obtener toda la información de una variable determinada. La aparición de datos censurados en estudios de confiabilidad es algo frecuente y pueden ser clasificados en diferentes tipos de acuerdo al fenómeno que generó dicha censura (Meeker & Escobar, 1998). Dentro de los tipos de censura la que más se presenta es la censura a derecha (Datta, 2005). Adicionalmente, investigaciones realizadas en Zhang & Jamshidian (2004) han propuesto diferentes alternativas computacionales en el trabajo de datos con doble censura. Un problema de interés se presenta cuando se tiene un grupo de dispositivos y sólo es posible observarlos en un punto del tiempo, un ejemplo concreto de esta situación son los bombillos de un tablero electrónico que no pueden permanecer bajo observación permanente, en tal situación se presenta una doble censura, es decir, algunos dispositivos fallan antes de x_0 (censurados a izquierda) y otros después de x_0 (censurados a derecha).

En diversas situaciones prácticas es de interés hallar la distribución del tiempo medio de falla para un conjunto de dispositivos teniendo únicamente la medición de un punto del tiempo. El principal objetivo de este artículo es realizar la estimación paramétrica del tiempo medio de falla cuándo en una situación determinada sólo se puede observar un punto del tiempo. En primera instancia, se hace una descripción del fenómeno a trabajar. Posteriormente, se aborda el fenómeno usando el modelo exponencial. Para esta distribución se encuentra de manera analítica el estimador de máxima verosimilitud (EMV) y los respectivos intervalos de confianza. El análisis anterior se hace bajo diferentes escenarios; el primero de ellos es cuando una proporción falla y otra aún está en funcionamiento $0 < n_0 < n$. Los dos casos adicionales se presentan cuando todos los dispositivos han fallado, $n = n_0$, o cuando ninguno ha fallado $n = n_1$. Posteriormente se presentan simulaciones de nivel real con el fin de comparar el intervalo exacto en caso de observar toda la muestra y el intervalo obtenido con doble censura.

Finalmente, se presenta una aplicación de la metodología desarrollada en datos de tubos de transmisión V805, que en su momento sirvieron como estándar para sistemas de radares. Se comparan los resultados que Davis (1952) presenta en su artículo frente a los propuestos por los autores en este artículo. Se concluye que la estimación del intervalo contiene los resultados hallados por Davis y resulta ser novedoso y útil dado que sólo se observa el conjunto de dispositivos una vez en el tiempo.

METODOLOGÍA

DESARROLLO ANALÍTICO EN EL CASO DE DOBLE CENSURA

Es bien conocido que en muchos problemas de ingeniería es de gran interés determinar la distribución del tiempo medio de falla de un dispositivo cuando sólo es posible observarlo en un punto del tiempo. Para tal fin, se realiza una prueba para n unidades con la característica que sólo es posible observar el resultado en un único punto del tiempo, x_0 . Se tendrá entonces que n_0 dispositivos ya fallaron y n_1 aún están en funcionamiento. Los datos obtenidos serán de la forma:

$$x_0^-, x_0^-, \dots, x_0^+, \dots, x_0^+$$

Se tiene que $n_0 x_0^-$ indica que ya fallaron pero no se registró el tiempo exacto y $n_1 x_0^+$ son las unidades que aún no han fallado.

Suponga que el tiempo de supervivencia tiene una distribución absolutamente continua $F(x; \theta)$, con densidad dada por $f(x; \theta)$.

Sea x_1, x_2, \dots, x_n una muestra aleatoria proveniente de una distribución exponencial, donde la verosimilitud es de la forma.

$$L(\theta|\text{datos}) = [F(x_0; \theta)]^{n_0} [1 - F(x_0; \theta)]^{n_1} \quad (1)$$

La verosimilitud será,

$$L(\lambda) = \left[1 - \exp\left(\frac{-x_0}{\lambda}\right)\right]^{n_0} \left[\exp\left(\frac{-x_0}{\lambda}\right)\right]^{n-n_0} \quad (2)$$

La log verosimilitud y la función Score son de la forma,

$$l(\lambda) = \log(L(\lambda)) = n_0 \log\left(1 - \exp\left(\frac{-x_0}{\lambda}\right)\right) - (n - n_0) \frac{x_0}{\lambda}$$

$$U = \frac{-dl(\lambda)}{d\lambda} = \frac{\frac{-n_0 x_0 \exp\left(\frac{-x_0}{\lambda}\right)}{\lambda^2}}{\left[1 - \exp\left(\frac{-x_0}{\lambda}\right)\right]} + \frac{(n - n_0)}{\lambda^2} \quad (3)$$

Resolviendo U se obtiene,

$$\hat{\lambda}(n_0) = \frac{x_0}{\log\left(\frac{n_0}{(n-n_0)+1}\right)} n_0 > 0 \quad (4)$$

Existen dos situaciones que se pueden presentar en este tipo de problema. La primera se da cuando se está observando y se conoce el número de fallas, de esta manera la información de Fisher sería.

$$I = E\left[-\frac{dU}{d\lambda}\right] = \frac{n_0 x_0}{\lambda^4 \left(\exp\left(\frac{x_0}{\lambda}\right) - 1\right)} x_0 - 2\lambda + \frac{x_0}{\left(\exp\left(\frac{x_0}{\lambda}\right) - 1\right)} + \frac{2n_1 \lambda}{n_0} \left(\exp\left(\frac{x_0}{\lambda}\right) - 1\right) \quad (5)$$

Por tanto, $V[\hat{\lambda}] = 1/I$. Además, si $\hat{\theta}$ es el estimador de máxima verosimilitud de θ , entonces bajo condiciones de regularidad $\hat{\theta} \sim N(\theta, I^{-1})$ (Kalbfleisch, 1985). Así un intervalo de confianza del 95% para la media cuando $n_0, n_1 \neq 0$ es.

$$\left(\lambda \pm 1,96 \sqrt{V[\hat{\lambda}]}\right) \quad (6)$$

La segunda situación se presenta cuando se asume que n_0 es una variable aleatoria, donde es posible resolver el problema condicionando n_0 , de esta forma es posible hallar $I(n_0)$.

$$I(n_0) = E\left[-\frac{dU}{d\lambda}\right] = \frac{n_0 x_0}{\lambda^4 \left(\exp\left(\frac{x_0}{\lambda}\right) - 1\right)} x_0 - 2\lambda + \frac{x_0}{\left(\exp\left(\frac{x_0}{\lambda}\right) - 1\right)} + \frac{2n_1 \lambda}{n_0} \left(\exp\left(\frac{x_0}{\lambda}\right) - 1\right) \quad (7)$$

El valor esperado es,

$$E[I(n_0)|_{0 < n_0 < n}] = \sum_{n_0=0}^{n-1} I(n_0)P(n_0) \quad (8)$$

Donde:

$$P(n_0) = \binom{n}{n_0} [F(x_0)]^{n_0} [1 - F(x_0)]^{n-n_0}$$

$$F(x_0) = 1 - \exp\left(-\frac{x_0}{\lambda}\right) \quad (9)$$

$$\sum_{j=1}^{n-1} P(X = j) = 1 - P(X = 0) - P(X = n) \quad (10)$$

Al calcular $E[I(n_0)|_{0 < n_0 < n}]$

$$E[I(n_0)|_{0 < n_0 < n}] = A + B \quad (11)$$

Con:

$$A = \frac{x_0 \left\{ \left[\exp\left(\frac{x_0}{\lambda}\right) - 1 \right] (x_0 - 2\lambda) - x_0 \right\}}{\hat{\lambda}^4 \left[\exp\left(\frac{x_0}{\lambda}\right) - 1 \right]} \sum_{n=1}^{n-1} \left[\frac{n \binom{n}{n_0} [F(x_0)]^{n_0} [1 - F(x_0)]^{n-n_0}}{1 - [1 - [F(x_0)]^n] - [F(x_0)]^n} \right]$$

$$B = \frac{2x_0}{\lambda^3} \sum_{n=1}^{n-1} \left[\frac{n \binom{n}{n_0} [F(x_0)]^{n_0} [1 - F(x_0)]^{n-n_0}}{1 - [1 - [F(x_0)]^n] - [F(x_0)]^n} \right] \quad (12)$$

Similar a la primera situación planteada se puede encontrar un intervalo de confianza del 95% para la media cuando $n_0, n_1 \neq 0$.

$$\left(\lambda \pm 1,96 \sqrt{V[\hat{\lambda}]} \right)$$

Si $n = n_0$ entonces se tendrá un intervalo unilateral $(0; \lambda_s)$, donde λ_s es el máximo valor posible de λ tal que $X_{(n)}$ es el n-ésimo estadístico de orden de una muestra de tamaño n que cumpla

$$P(X_{(n)} \leq x_0) \geq 0.05$$

La función densidad de probabilidad de $X_{(n)}$ está dada por

$$g_n(x_{(n)}) = n \left(F(x_{(n)}) \right)^{n-1} f(x_{(n)}) \quad (13)$$

El intervalo unilateral $(0; \lambda_s)$ puede considerarse como el conjunto de λ^-s que puede producir una muestra donde todos los elementos de la muestra de tamaño n fallen antes del punto x_0 con una probabilidad mayor o igual a 0.05.

$$P(X_{(n)} \leq x_0) = G_n(x_0) = (F(x_0))^n = \left(1 - \exp\left(-\frac{x_0}{\lambda}\right) \right)^n \geq 0.05 \quad (14)$$

Donde $G_n(x_0)$ es la función de distribución acumulada de $X_{(n)}$. Por tanto,

$$\lambda_S = -\frac{x_0}{\log\left(1-0.05^{\frac{1}{n}}\right)} \tag{15}$$

Finalmente, si $n = n_1$ entonces se tendrá un intervalo unilateral $(\lambda_l; +\infty)$, donde λ_l es el mínimo valor posible de λ tal que $X_{(1)}$ es el menor estadístico de orden que cumpla.

$$P(X_{(1)} \geq x_0) \geq 0.05$$

La función de densidad de probabilidad de $X_{(1)}$ está dada por:

$$g_1(x_{(1)}) = n[1 - F(x_{(1)})]^{n-1} f(x_{(1)}) \tag{16}$$

El intervalo unilateral $(\lambda_l; +\infty)$ puede ser considerado como el conjunto de λ^+ s que pueden producir una muestra donde todos los elementos de la muestra de tamaño n no fallen antes del punto x_0 con una probabilidad mayor o igual a 0.05, o lo que es lo mismo, que el menor tiempo de falla en la muestra sea mayor que x_0 .

$$P(X_{(1)} \geq x_0) = \exp\left(-\frac{nx_0}{\lambda}\right) \geq 0.05 \tag{17}$$

De este proceso resulta:

$$\lambda_l = -\frac{nx_0}{\log(0.05)} \tag{18}$$

Este valor es el mínimo valor de la media λ que produciría una muestra de n elementos que no fallarán hasta el punto de observación x_0 con una probabilidad de 0.95.

RESULTADOS

Simulaciones de nivel real

Se realizaron simulaciones con el fin de comparar el intervalo exacto que se calcularía en caso de tener la muestra completamente observada, lo que sería ideal, y el intervalo que se calcularía en caso de tener sólo datos censurados en x_0 , tanto a izquierda como a derecha. La comparación se realiza analizando tanto la longitud promedio de los intervalos como el nivel real de confianza. Se realizaron 10.000 simulaciones para diferentes tamaños muestrales y diferentes puntos del tiempo, con un nivel nominal de confianza del 95%.

Si se tuvieran las observaciones completas, es decir, $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ se podrían emplear para construir el intervalo de confianza para la media de la exponencial. Se sabe que $T = \sum_{i=1}^n X_i \sim \text{Gamma}(n\lambda)$, entonces $\frac{2T}{\lambda} \sim X_{2n}^2$ (Kalbfleisch, 1985). El intervalo de confianza exacto de nivel $(1 - \alpha)100\%$ presentado por Meeker & Escobar (1998) y Lawless (1982) es:

$$\frac{2 \sum_{i=1}^n X_i}{X_{2n, 1-\alpha/2}^2} \leq \frac{2 \sum_{i=1}^n X_i}{X_{2n, \alpha/2}^2} \tag{19}$$

El procedimiento usado para las simulaciones se presenta como sigue:

1. Elección de los tamaños muestrales. Se consideran tamaños muestrales de 10, 20, 50, 100, 200, 500.
2. Elección de los puntos del tiempo donde se observa la muestra. Se observan en los puntos 0,25; 0,5; 1; 2,3, 4.
3. Se hallan todas las combinaciones entre puntos del tiempo y tamaños muestrales.
4. Cálculo de la longitud promedio de cada intervalo y su respectivo nivel real.

Tabla 1. Resultados de simulación de nivel real: tiempos completos vs tiempos censurados.

RESULTADOS DE SIMULACIONES							
	x_0	λ	n	Doble censura		Tiempos completos	
				Media longitud	Nivel alcanzado	Media longitud	Nivel real
1	0,25	1	10	3,144	0,823	1,466	0,955
2	0,25	1	20	2,717	0,944	0,961	0,950
3	0,25	1	50	1,416	0,931	0,576	0,950
4	0,25	1	100	0,905	0,933	0,399	0,949
5	0,25	1	200	0,613	0,952	0,280	0,949
6	0,25	1	500	0,380	0,951	0,176	0,948
7	0,5	1	10	2,863	0,845	1,5	0,952
8	0,5	1	20	1,761	0,949	0,964	0,951
9	0,5	1	50	0,960	0,953	0,575	0,952
10	0,5	1	100	0,655	0,948	0,400	0,951
11	0,5	1	200	0,454	0,946	0,280	0,952
12	0,5	1	500	0,284	0,948	0,176	0,951
13	1	1	10	1,987	0,941	1,513	0,956
14	1	1	20	1,255	0,913	0,964	0,953
15	1	1	50	0,753	0,959	0,576	0,948
16	1	1	100	0,522	0,951	0,399	0,950
17	1	1	200	0,366	0,955	0,280	0,949
18	1	1	500	0,231	0,950	0,176	0,950
19	2	1	10	1,860	1	1,638	0,968
20	2	1	20	1,165	1	0,978	0,960
21	2	1	50	0,709	0,914	0,576	0,950
22	2	1	100	0,498	0,962	0,399	0,952
23	2	1	200	0,352	0,954	0,280	0,947
24	2	1	500	0,222	0,955	0,176	0,950
25	3	1	10	2,302	1	1,852	0,939
26	3	1	20	1,402	0,999	1,049	0,961
27	3	1	50	0,826	0,997	0,583	0,959
28	3	1	100	0,576	0,957	0,399	0,953

30	3	1	500	0,256	0,944	0,176	0,949
31	4	1	10	2,891	1	2,045	0,878
32	4	1	20	1,758	0,999	1,126	0,926
33	4	1	50	1,035	0,996	0,604	0,953
34	4	1	100	0,730	0,988	0,405	0,956
35	4	1	200	0,517	0,987	0,280	0,955
36	4	1	500	0,323	0,937	0,176	0,952

Análisis de resultados de simulación

De la tabla 1 se observa que en términos generales los resultados de las simulaciones muestran que la longitud promedio de los intervalos en caso de datos censurados son mayores que la longitud promedio de los intervalos exactos. Sin embargo, es aceptable dado que se está tomando información en un único punto del tiempo frente a la observación de toda la muestra.

En la figura 1 se compara la longitud promedio de intervalos aproximados e intervalos exactos para los diferentes puntos del tiempo, x_0 , en los que se realizaron las simulaciones.

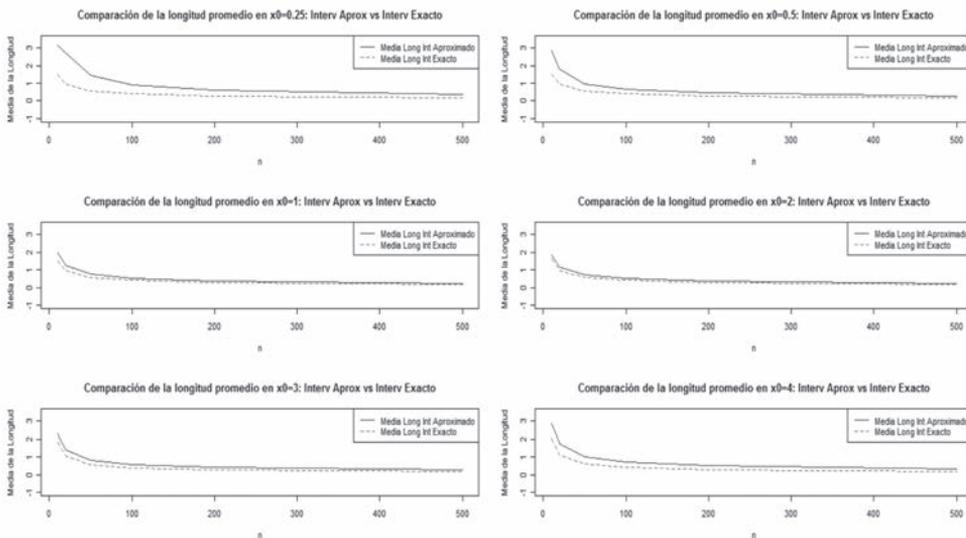


Figura 1. Comparación de las longitudes medias de los intervalos aproximados para diferentes x_0 .

Los resultados obtenidos en las simulaciones muestran que las longitudes promedio de los intervalos en el caso de datos censurados son mayores que la longitud media de los intervalos exactos. Esto se evidencia para el caso en que el punto de observación sea menor al valor medio de la distribución, por el contrario, cuando la media es inferior al punto de corte x_0 dicha diferencia se hace más pequeña.

Aplicación: Tubos V805

Davis (1952), consideró un modelo exponencial para la duración de tubos de transmisión V805, los cuales fueron en el pasado el estándar en sistemas de radares de aviones. El primer paso es realizar una prueba de bondad de ajuste, dada la naturaleza de la información presentada se realiza una prueba chi-cuadrado en la que se valida que efectivamente los datos siguen una distribución exponencial con media 179. A continuación se presenta el cálculo del tiempo medio de falla con su respectivo intervalo de confianza del 95 % tomando la aproximación clásica exponencial, lo anterior con el fin de compararlos con los resultados obtenidos por Davis.

De 903 tubos analizados, a las 40 horas de funcionamiento 166 habían fallado y los demás seguían en funcionamiento. Bajo este panorama se tiene que $x_0 = 40$ y $n = 166$. De esta manera es posible obtener el estimador puntual y el intervalo de confianza del 95% para el tiempo medio de falla de los tubos V805.

Bajo el modelo exponencial y empleando los resultados del análisis realizado anteriormente se tiene que los resultados son:

Tiempo medio de falla de tubos V805: 196.9137 horas

Intervalo de confianza asintótico del 95% para el tiempo medio de falla tubos V805:
(169,3183; 224,5091)

Comparando los resultados obtenidos se encuentra que el intervalo de confianza del 95% construido con la metodología propuesta en este trabajo contiene el tiempo medio de falla hallado por (Davis, 1952), el cual fue 179 horas y muestra que es posible encontrar un intervalo de confianza para el tiempo medio de falla en situaciones donde sólo es posible observar un punto del tiempo.

DISCUSIÓN

La metodología presentada permite resolver problemas de tipo práctico en los que es costoso e improbable observar durante un periodo de tiempo un fenómeno determinado.

La metodología planteada en el artículo es fácil de usar e interpretar. De gran aplicación a problemas en los que no se pueda mantener un conjunto de dispositivos en observación durante periodos de tiempo determinados.

CONCLUSIONES

Las simulaciones realizadas en la aproximación clásica al problema permiten concluir que las longitudes de los intervalos exactos contra los intervalos aproximados por grupo se encuentra que en términos generales que las longitudes de los intervalos exactos son menores a las de los intervalos aproximados. Y conforme n aumenta la longitud del intervalo de confianza tiende a ser más estable y cercano a cero.

Al comparar las estimaciones obtenidas se concluye que es posible asumir el riesgo de hacer estimación teniendo sólo en cuenta un punto del tiempo.

En problemas donde se tiene un conjunto de dispositivos y sólo es posible observar un punto del tiempo este tipo de metodologías son recomendables puesto que la pérdida no es considerablemente alta.

REFERENCIAS

DATTA, S. Estimating the mean life time using right censored data. *Statistical Methodology*, 2005, vol. 2, no. 1, pp. 65-69. DOI 10.1016/j.stamet.2004.11.003. Elsevier BV

DAVIS, D. An Analysis of Some Failure Data. *Journal of the American Statistical Association*, 1952, vol. 47, no. 258, pp. 113-150. DOI 10.1080/01621459.1952.10501160. Informa UK Limited

KALBFLEISCH, J. *Probability and Statistical Inference*. 1985, New York, NY: Springer New York.

LAGAKOS, S. General Right Censoring and Its Impact on the Analysis of Survival Data. *Biometrics*, 1979, vol. 35, no. 1, pp. 139. DOI 10.2307/2529941. JSTOR

LAWLESS, J. *Statistical models and methods for lifetime data*. 1982, New York: Wiley.

ZHANG, Y. and JAMSHIDIAN, M. On Algorithms for the Nonparametric Maximum Likelihood Estimator of the Failure Function With Censored Data. *Journal of Computational and Graphical Statistics*, 2004, vol. 13, no. 1, pp. 123-140. DOI 10.1198/1061860043038. Informa UK Limited

