

Inferencia estadística para mezclas finitas de distribuciones con censura e identificación parcial usando variables latentes

Alberto Contreras-Cristán y Federico O'Reilly

Instituto de Investigaciones en
Matemáticas Aplicadas y en Sistemas
Universidad Nacional Autónoma de México
alberto@sigma.iimas.unam.mx

1. Introducción

VARIABLES aleatorias cuya distribución está descrita por una mezcla con un número finito de componentes pueden encontrarse en varios ejemplos de aplicaciones de la ciencia, en particular en el estudio de tiempos de vida o en el funcionamiento de aparatos, en teoría de *supervivencia* ó *confiabilidad*. Al inicio de su uso, ciertos dispositivos electrónicos presentan tasas de falla que no son pequeñas y, a lo largo de su vida, estas tasas de falla decrecen para luego incrementarse con la edad del dispositivo, de forma que la idea de un modelo de mezcla para el tiempo de vida del dispositivo resulta natural. Por otra parte, cuando el dispositivo está formado por componentes, asumiendo que al fallar el dispositivo se puede averiguar cuál de estas componentes ha fallado, se puede asociar la componente (ó tipo de falla) con un término en la mezcla que define la distribución de los tiempos de vida. Preguntas de interés en relación a esta situación son: ¿cuál es el porcentaje de dispositivos que fallan debido a un tipo de falla específico? ó ¿cuál es la distribución de cada tipo de falla?

Supóngase que observamos datos que provienen de una mezcla de distribuciones con k componentes. Para cada $j = 1, \dots, k$, sea $F_j(\cdot; \theta_j)$ la función de distribución correspondiente a la j -ésima componente en el modelo de mezcla (la distribución) para las observaciones, el cual

está dado por

$$G(x; \boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{j=1}^k \pi_j F_j(x; \theta_j), \quad (1)$$

donde los parámetros del modelo $\boldsymbol{\theta}' = (\theta_1, \theta_2, \dots, \theta_k)$ y $\boldsymbol{\pi}' = (\pi_1, \pi_2, \dots, \pi_k)$ son desconocidos y tales que $\boldsymbol{\theta}' \in \mathbb{R}^k$, $\boldsymbol{\pi}' \in [0, 1]^k$ y $\sum_{j=1}^k \pi_j = 1$. Denotaremos por $f_j(\cdot; \theta_j)$ a la función de densidad correspondiente a $F_j(\cdot; \theta_j)$, $j = 1, 2, \dots, k$.

Por razones económicas ó de logística, en estudios de tiempos de falla suele suceder que hay un límite de tiempo después del cual el estudio ya no continúa, supongamos que este valor límite o umbral se denota por C . Alternativamente, el final del estudio puede establecerse cuando ocurre un número de fallas, es decir que cuando se completa un número predeterminado de fallas se detiene el estudio. En el primero de estos casos, se dice que las observaciones tienen censura por la derecha en C , ya que para un tiempo de falla que excede el valor de C , como el estudio ya no continúa entonces no necesariamente se conoce cuál fue el valor del tiempo de falla.

2. Un problema de tiempos de falla con censura e identificación parcial

Adicionalmente al modelo (1), asumiremos que cada observación sigue un esquema de censura en el cual si el dato excede un umbral C , entonces no podemos saber ni cuánto vale el dato ni de qué componente de la mezcla (1) proviene, solo obtenemos la información de que el dato excede el valor de C . Por otra parte, si el dato no excede el valor de C , entonces sí podemos registrar su valor y además podemos identificar la componente de la mezcla (1) de la cual el dato proviene.

En el supuesto de que después de hacer N observaciones independientes siguiendo el esquema descrito arriba, resulta que $N - r$ tiempos de falla fueron mayores que C y que, de los r que fueron menores o iguales a C se identificaron como: r_1 provenientes de la primera componente de la mezcla, r_2 de la segunda componente de la mezcla, y así hasta r_k de la k -ésima componente, entonces la información obtenida de estas N observaciones puede resumirse como las k colecciones $\{x_{1i}\}_{i=1}^{r_1}$, $\{x_{2i}\}_{i=1}^{r_2}$, \dots , $\{x_{ki}\}_{i=1}^{r_k}$, así como el conteo $N - r \equiv N - \sum_{j=1}^k r_j$, donde para cada $j = 1, \dots, k$, x_{j1}, \dots, x_{jr_j} , denotan a las r_j observaciones menores o iguales que C y que provienen de la componente j en la mezcla.

Para llevar a cabo la estimación de los parámetros en (1) vía máxima verosimilitud, necesitamos escribir la densidad conjunta de datos con la estructura planteada, para obtener intuición, comencemos por escribir

el caso $k = 1$ que corresponde a un solo término en (1) con $\pi_1 = 1$, $r_1 = r$ y $G(x; \theta) = F_1(x; \theta)$. Consideremos el evento

$$A_1 = [X_1 \leq C; X_2 \leq C; \dots; X_r \leq C] \cap [N - r \text{ observaciones son } > C],$$

el cual tiene probabilidad $P(A_1) = \frac{N!}{(N-r)!r!} [F_1(C; \theta)]^r [1 - F_1(C; \theta)]^{N-r}$, la densidad conjunta de X_1, \dots, X_r , condicional a A_1 y evaluada en las observaciones $\{x_i\}_{i=1}^r$ es

$$f_{X_1, \dots, X_r}(x_1; \dots; x_r | A_1) = \prod_{i=1}^r f_{X_i}(x_i | X_i \leq C) = \prod_{i=1}^r \left\{ \frac{f_1(x_i; \theta)}{F_1(C; \theta)} \right\}.$$

De manera que la verosimilitud para el caso $k = 1$ es

$$\begin{aligned} L(\theta) &= f_{X_1, \dots, X_r}(x_1; \dots; x_r | A_1) P(A_1) \\ &= C_1 \left\{ \prod_{i=1}^r f_1(x_i; \theta) \right\} [1 - F_1(C; \theta)]^{N-r}, \end{aligned}$$

donde $C_1 = \frac{N!}{(N-r)!r!}$ no depende de θ . Para $k = 2$ hay 2 términos en (1), donde $\pi_2 = 1 - \pi_1$, $r = r_1 + r_2$ y $\theta' = (\theta_1, \theta_2)$, de manera que $G(C; \theta; \pi) = \sum_{j=1}^2 \pi_j F_j(C; \theta_j)$, ahora condicionaremos en el evento

$$\begin{aligned} A_2 &= [X_{1,1} \leq C; \dots; X_{1,r_1} \leq C] \cap [X_{2,1} \leq C; \dots; X_{2,r_2} \leq C] \\ &\quad \cap [N - r \text{ observaciones son } > C], \end{aligned}$$

tal que

$$P(A_2) = \frac{N!}{(N-r)!r_1!r_2!} [\pi_1 F_1(C; \theta_1)]^{r_1} [\pi_2 F_2(C; \theta_2)]^{r_2} [1 - G(C; \theta; \pi)]^{N-r}.$$

La densidad condicional al evento A_2 , para $\{X_{1i}\}_{i=1}^{r_1}$, $\{X_{2i}\}_{i=1}^{r_2}$ es

$$\begin{aligned} f_{\{X_{1i}\}_{i=1}^{r_1}, \{X_{2i}\}_{i=1}^{r_2}}(x_{11}; \dots; x_{1r_1}; x_{21}; \dots; x_{2r_2} | A_2) &= \prod_{j=1}^2 \prod_{i=1}^{r_j} f_{X_{ji}}(x_{ji} | X_{ji} \leq C) \\ &= \prod_{j=1}^2 \prod_{i=1}^{r_j} \frac{f_j(x_{ji}; \theta_j)}{F_j(C; \theta_j)}. \end{aligned} \tag{2}$$

Multiplicando (2) por $P(A_2)$ obtenemos la verosimilitud para el caso $k = 2$

$$L(\theta, \pi) = C_2 \left\{ \prod_{j=1}^2 \prod_{i=1}^{r_j} \frac{f_j(x_{ji}; \theta_j)}{F_j(C; \theta_j)} \right\} \prod_{j=1}^2 \{ \pi_j F_j(C; \theta_j) \}^{r_j} \times \{ 1 - G(C; \theta; \pi) \}^{N-r},$$

donde $C_2 = \frac{N!}{(N-r)!r_1!r_2!}$ no depende de $\boldsymbol{\theta}$. En general, para k componentes en (1), la función de verosimilitud está dada por

$$L(\boldsymbol{\theta}, \boldsymbol{\pi}; \mathbf{x}) = C_k \left\{ \prod_{j=1}^k \prod_{i=1}^{r_j} \frac{f_j(x_{ji}; \theta_j)}{F_j(C; \theta_j)} \right\} \prod_{j=1}^k \{\pi_j F_j(C; \theta_j)\}^{r_j} \times \{1 - G(C; \boldsymbol{\theta}, \boldsymbol{\pi})\}^{N-r}, \quad (3)$$

donde $C_k = \frac{N!}{(N-r)!r_1!r_2!\dots r_k!}$ y $G(C; \boldsymbol{\theta}; \boldsymbol{\pi}) = \sum_{j=1}^k \pi_j F_j(C; \theta_j)$.

Mendenhall y Hader (1958) estudiaron los estimadores de máxima verosimilitud provenientes de (3), asumiendo $k = 2$ y que las distribuciones $F_j, j = 1, 2$, son exponenciales. Para su análisis, utilizan datos de tiempos de falla de dispositivos de telecomunicaciones utilizados por una línea aérea comercial, si la falla acontece antes de 630 ($= C$) horas de uso, el dispositivo es retirado del avión, pero si el dispositivo no ha fallado al llegar a las 630 horas de todas formas es reemplazado, este es el plazo establecido por las políticas de mantenimiento de la compañía aérea. Una vez que acontece, la falla se puede catalogar en uno de dos tipos: ($j = 1$) aquellos que corresponden a fallas confirmadas y ($j = 2$) fallas no confirmadas, esta última clasificación surge porque en algunos casos, cuando el dispositivo llegaba al centro de mantenimiento después de ser reemplazado, resultaba que mostraba un funcionamiento satisfactorio.

Díaz-Francés (1998) propuso una solución a este problema utilizando verosimilitudes perfil (Venzon y Moolgavkar, 1988; McCullagh y Tibshirani, 1990). Supóngase que para una muestra aleatoria de variables $\mathbf{y}_n = (y_1, \dots, y_n)$, independientes y todas con la misma distribución, la función de verosimilitud está dada por $L(\Theta, \mathbf{y})$, además imaginemos que el parámetro Θ se puede particionar como $\Theta = (\psi, \lambda)$, donde nuestro interés primario está en ψ . Entonces denotando por $\hat{\psi}_\lambda$ al estimador de máxima verosimilitud de ψ considerando λ fijo (y en forma análoga al estimador de máxima verosimilitud de λ considerando ψ fijo por $\hat{\lambda}_\psi$), la verosimilitud perfil para λ es

$$L^p(\lambda) = L(\hat{\psi}_\lambda, \lambda, \mathbf{y}).$$

Denotando por $\hat{\lambda}$ al máximo de esta función con respecto a λ , el estimador de máxima verosimilitud para ψ obtenido con este método es $\hat{\psi}_{\hat{\lambda}}$. Para su análisis, Díaz-Francés considera un modelo Weibull con parámetros de forma α_j y escala β_j para $F_j, j = 1, 2$,

$$\tilde{F}_j(y|\alpha_j, \beta_j) = 1 - \exp\{-(y/\beta_j)^{\alpha_j}\},$$

donde para facilitar cálculos, se usan la transformaciones $X = \log(Y/C)$, $\mu_j = \log(\beta_j/C)$ y $\sigma_j = 1/\alpha_j$, de forma que

$$F_j(x|\mu_j, \sigma_j) = 1 - \exp\{-e^{(x-\mu_j)/\sigma_j}\} \quad (4)$$

es una distribución de valores extremos con parámetro de localización μ_j y parámetro de escala σ_j . Después de estas transformaciones, el nuevo umbral de censura para los datos es $C_0 = 0$. En la sección que sigue, propondremos el uso de variables latentes para estudiar la estimación de los parámetros del modelo, para lo cual usaremos la transformación (4).

3. Análisis de Verosimilitud perfil vía variables latentes

3.1 Variables latentes

Para cada $j = 1, 2, \dots, k$, supóngase que r_j^* es el número de datos que no observamos (porque exceden el valor de C) y que provienen de la distribución F_j , cada r_j^* es un entero mayor o igual a 0 y además $\sum_{j=1}^k r_j^* = N - r$. Las variables latentes (no observadas) $r_1^*, r_2^*, \dots, r_k^*$ nos permitirán hacer un análisis alternativo. Si los valores de r_j^* ; $j = 1, 2, \dots, k$ fueran conocidos, entonces los pesos en la mezcla (1) ya no juegan un papel en el modelo, puesto que se conoce de qué componente de (1) vienen los datos (aunque estos excedan el valor de C), de esta forma, si $B_j = [X_{j1} \leq C; \dots; X_{jr_j} \leq C] \cap [r_j^* \text{ observaciones son } > C]$, tenemos

$$P(B_j) = \binom{r_j + r_j^*}{r_j} [F_j(C; \theta_j)]^{r_j} \{1 - F_j(C; \theta_j)\}^{r_j^*},$$

de donde la contribución a la verosimilitud $L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{r}^*)$ correspondiente a la j -ésima componente en (1) es

$$\begin{aligned} L_j(\theta_j; \mathbf{x}_j, r_j^*) &= f_{X_{j1}, \dots, X_{jr_j}}(x_{j1}; \dots; x_{jr_j} | B_j) P(B_j) \\ &= \left\{ \prod_{i=1}^{r_j} \frac{f_j(x_{ji}; \theta_j)}{F_j(C; \theta_j)} \right\} P(B_j) \\ &= \left\{ \prod_{i=1}^{r_j} f_j(x_{ji}; \theta_j) \right\} \binom{r_j + r_j^*}{r_j} \{1 - F_j(C; \theta_j)\}^{r_j^*}. \quad (5) \end{aligned}$$

De manera que la función de verosimilitud $L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{r}^*)$ toma la forma

$$L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{r}^*) = \prod_{j=1}^k L_j(\theta_j; \mathbf{x}_j, r_j^*). \quad (6)$$

Es decir que al asumir que conocemos $r_1^*, r_2^*, \dots, r_k^*$, podemos factorizar la función de verosimilitud para θ como un producto de funciones de verosimilitud marginales, donde cada L_j contiene toda la información necesaria para poder hacer inferencias acerca de la componente θ_j de θ . Para cada valor de fijo de r_j^* sea $\hat{\theta}_j(r_j^*)$ el estimador de máxima verosimilitud para θ_j . Entonces podemos usar la función de verosimilitud perfil para $r_1^*, r_2^*, \dots, r_k^*$, dada por $L^p(r_1^*, \dots, r_k^*) = \prod_{j=1}^k L_j(\hat{\theta}_j(r_j^*); \mathbf{x}_j, r_j^*)$, para encontrar los valores \hat{r}_j^* , $j = 1, 2, \dots, k$, que la maximizan, una vez hecho esto, los estimadores de máxima verosimilitud para $\theta_1, \dots, \theta_k$ están dados por

$$\hat{\theta}_j(\hat{r}_j^*), \quad j = 1, 2, \dots, k. \quad (7)$$

Llamaremos a este método de estimación *Análisis de Verosimilitud Perfil Usando Variables Latentes* (LVPLA) y, para ilustrar su uso, en la sección 3.3 usaremos los datos de dispositivos en Mendenhall y Hader (1958) con el modelo (4).

3.2 Aproximación normal para la función de verosimilitud

Una forma de construir intervalos de confianza para los parámetros de un modelo, consiste en utilizar una aproximación Normal para la función de verosimilitud del parámetro. El método se basa en encontrar la densidad Normal que mejor aproxima a la función de verosimilitud (véase Kalbfleisch, 1985, sección 9.7). Sean $L^p(\lambda)$ la verosimilitud perfil para λ y $\hat{\lambda}$ el valor de λ que maximiza a $L^p(\lambda)$. Si $r(\lambda) = \log\left(\frac{L^p(\lambda)}{L^p(\hat{\lambda})}\right)$, entonces al calcular la expansión de Taylor de segundo orden para $\log(L^p(\lambda))$, alrededor de $\hat{\lambda}$, obtenemos que $r(\lambda)$ es aproximadamente $-\frac{1}{2}(\lambda - \hat{\lambda})^2 I(\hat{\lambda})$, donde $I(\lambda) = -\frac{d^2 \log(L^p(\lambda))}{d\lambda^2}$ es la función de información. Esta aproximación se considera adecuada cuando el tamaño de muestra es grande, ya que si esto sucede resulta suficiente considerar la expansión de Taylor solo hasta un segundo orden. La cantidad $I(\hat{\lambda})$ se conoce como *la información observada de Fisher*, Efron y Hinkley (1978) sugieren usar $I(\hat{\lambda})$ en lugar de la información (esperada) de Fisher cuando se usan aproximaciones Normales. En el caso en que λ es una cantidad vectorial, digamos $\lambda' = (\lambda_1, \dots, \lambda_m)$, la información observada de Fisher es una matriz de segundas derivadas con entrada (i, j) dada por $-\frac{\partial^2 \log(L^p(\lambda))}{\partial \lambda_i \partial \lambda_j} \Big|_{\lambda=\hat{\lambda}}$, $i, j \in \{1, \dots, m\}$.

Regresando a la construcción de intervalos de la discusión en el párrafo anterior, si $N(\lambda; \hat{\lambda}, I^{-1}(\hat{\lambda}))$ denota la densidad Normal con media $\hat{\lambda}$ y varianza $I^{-1}(\hat{\lambda})$, evaluada en λ , entonces para tamaños de muestra

grandes se tiene la aproximación

$$e^{r(\lambda)} = \frac{L^p(\lambda)}{L^p(\hat{\lambda})} \approx \frac{N(\lambda; \hat{\lambda}, I^{-1}(\hat{\lambda}))}{N(\hat{\lambda}; \hat{\lambda}, I^{-1}(\hat{\lambda}))} \equiv e^{r_N(\lambda)}.$$

Para $p \in (0, 1)$, el intervalo aproximado para λ , a nivel $100p\%$, es el conjunto de valores de λ tales que $r_N(\lambda) \geq \log(p)$, condición que nos lleva a que este intervalo es

$$\left(\hat{\lambda} - \sqrt{(-2 \log(p))I^{-1}(\hat{\lambda})}, \hat{\lambda} + \sqrt{(-2 \log(p))I^{-1}(\hat{\lambda})} \right).$$

3.3 El caso de datos de dispositivos transmisores

Habiendo llevado a cabo las transformaciones y reparametrizaciones descritas al final de la sección 2, obtenemos que

$$f_j(x_{ji}|\theta_j) = e^{-\varphi_j + (x_{ji} - \mu_j)e^{-\varphi_j}} \exp \left\{ -e^{(x_{ji} - \mu_j)e^{-\varphi_j}} \right\}$$

y

$$F_j(x_{ji}|\theta_j) = 1 - \exp \left\{ -e^{(x_{ji} - \mu_j)e^{-\varphi_j}} \right\},$$

donde $\theta_j = (\mu_j, \varphi_j)'$. Para los datos de dispositivos de telecomunicaciones, tenemos $k = 2$, $N = 369$, $r_1 = 107$, $r_2 = 218$ y estamos considerando dos variables latentes r_1^* y r_2^* , pero como $r_2^* = N - r - r_1^*$, entonces nos concentraremos únicamente en r_1^* . Al implementar el (LVPLA) en este conjunto de datos, la Figura 1 nos muestra valores numéricos (estandarizados) de la función de verosimilitud perfil para r_1^* , como podemos ver el máximo se da en $r_1^* = 2$, de forma que asumiremos que $r_1^* = \hat{r}_1^* = 2$ para continuar el análisis.

Notemos que el uso del LVPLA introduce simplificaciones en la forma funcional de la verosimilitud (ecuaciones 5 y 6) que permiten evaluar la función de verosimilitud perfil para μ_1 maximizando $L_1(\mu_1, \varphi_1; \mathbf{x}_1, r_1^*)$ únicamente con respecto a φ_1 . En contraste, sin usar las variables latentes r_1^* y r_2^* , la función de verosimilitud perfil para μ_1 se hubiera tenido que obtener a partir de $L(\mu_1, \mu_2, \varphi_1, \varphi_2, \pi_1; \mathbf{x})$ en (3), maximizando sobre cuatro parámetros $\mu_2, \varphi_1, \varphi_2$ y π_1 . El LVPLA también tiene ventajas computacionales al usar la aproximación Normal descrita en la sección 3.2 para el cálculo de intervalos de confianza correspondientes a los parámetros. En el caso del análisis para μ_1 , este método nos permite calcular la varianza asintótica utilizando la información observada de Fisher calculada a partir de $L_1(\mu_1, \varphi_1; \mathbf{x}_1, r_1^*)$. Sin usar el LVPLA, para obtener la aproximación Normal se requeriría la información observada de Fisher calculada a partir de $L(\mu_1, \mu_2, \varphi_1, \varphi_2, \pi_1; \mathbf{x})$ en (3). Las anteriores observaciones sobre las ventajas del uso del LVPLA, se vuelven más significativas cuando el número de términos k en la mezcla

(1) crece. Por otra parte, el LVPLA se puede estudiar para otras familias de distribuciones con parámetros de localización y escala diferentes al modelo Weibull, por ejemplo, usando resultados como en Anaya y O'Reilly (2000). En Campos et al. (2010) se estudia el LVPLA para mezclas de distribuciones Normales.

Para los datos de transmisores de Mendenhall y Hader (1958), la tabla 1 muestra los intervalos (con nivel 95 % de confianza) obtenidos para los parámetros μ_1 , μ_2 , φ_1 y φ_2 al usar la aproximación Normal. También en la tabla 1, aparecen intervalos de confianza (con nivel 95 %) obtenidos con un segundo método de estimación conocido como el algoritmo

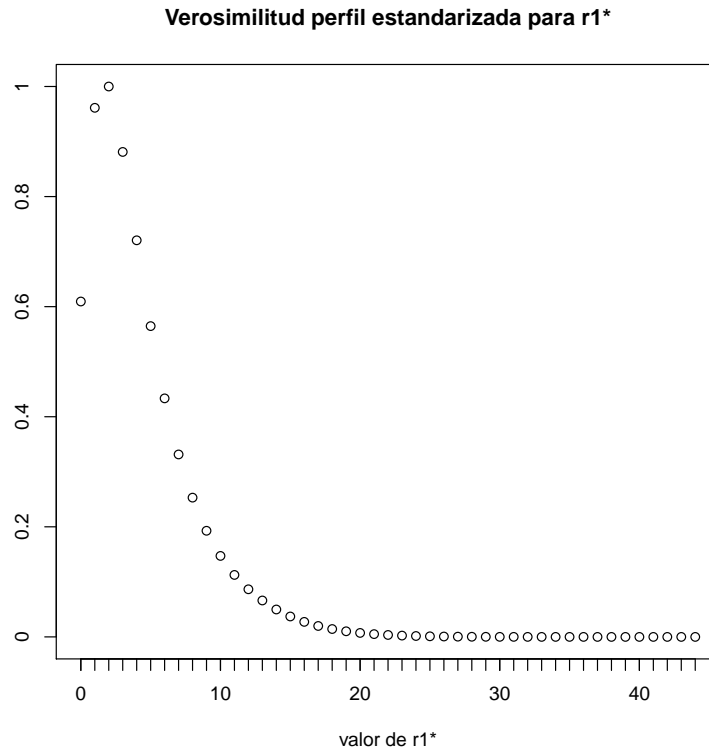


Figura 1. Función de verosimilitud perfil (estandarizada) para r_1^* .

Límites superiores e inferiores para los intervalos a nivel 95 %.

	EM	LVPLA
μ_1	(-1.24,-0.85)	(-1.21,-0.91)
μ_2	(-0.70,-0.45)	(-0.71,-0.45)
φ_1	(-0.42, -0.05)	(-0.41,-0.11)
φ_2	(-0.23,-0.01)	(-0.22,-0.01)

Cuadro 1. Resultados para los métodos EM y LVPLA.

	Ξ_1	Ξ_2	Ξ_3
μ_1	-1.0451	-0.9836	-1.1374
μ_2	-0.5781	-0.5163	-0.6353
φ_1	-0.2362	-0.1776	-0.3273
φ_2	-0.1180	-0.0607	-0.1714
π_1	0.2954	0.26	0.33

Cuadro 2. Valor de los parametros para simulación ($\pi_2 = 1 - \pi_1$).

EM (Tanner, 1996, capítulo 4). En Contreras-Cristán et al. (2003), así como en Campos et al. (2010), se compara el LVPLA con el algoritmo EM y con un método Bayesiano (BALV), para diferentes valores de k y tipos de censura.

En cuanto a la inferencia para los pesos de la mezcla, aunque estos no aparecen en forma explícita cuando se lleva a cabo el LVPLA, podemos estimarlos notando que como un total de $r_1 = 107$ observaciones fueron identificadas como provenientes de la primera población y además nosotros obtuvimos del LVPLA, que $r_1^* = 2$ hubieran contado como observaciones de la primera población que fueron censuradas, se concluye que habría 109 observaciones de la población 1. Como el número total de observaciones en la muestra es 369, entonces una estimación de π_1 es $\hat{\pi}_1 = 109/369 = 0.29$, que a su vez nos dice que $\hat{\pi}_2 = 0.71$.

4. Estudio comparativo de coberturas

Se considera que el algoritmo EM lleva a cabo un cálculo numérico exacto de estimadores al maximizar directamente la verosimilitud (3) con respecto a (θ, π) , mientras que el LVPLA aproxima estos estimadores al maximizar la *verosimilitud aumentada* (6) con respecto a (θ, \mathbf{r}^*) , no obstante esta aproximación resulta adecuada, pero sobre todo eficiente, ya que como hemos comentado se simplifica la forma funcional de (3) y en el caso de que k sea mayor a 2, esta simplificación puede proveernos con procedimientos numéricos económicos. Para fines de comparación, se llevó a cabo un estudio de coberturas en el cual: para cada valor de los parámetros en la Tabla 2, se simulan 100 muestras con el esquema de censura descrito en la sección 2, para cada una de estas muestras se calculan los intervalos usando el LVPLA y el algoritmo EM, se utilizan los 100 intervalos obtenidos con cada método para estimar la cobertura (la proporción de intervalos simulados que contiene a los valores de los parámetros en la Tabla 2). Los resultados del estudio están en las Tablas 3 (LVPLA) y 4 (algoritmo EM).

	Ξ_1		Ξ_2		Ξ_3	
	cober.	long. prom.	cober.	long. prom.	cober.	long. prom.
μ_1	0.85	(0.3100)	0.86	(0.3500)	0.92	(0.2700)
μ_2	0.93	(0.2400)	0.94	(0.2500)	1.00	(0.2200)
φ_1	0.87	(0.3170)	1.00	(0.3500)	0.91	(0.2900)
φ_2	0.99	(0.2400)	0.98	(0.2400)	1.00	(0.2400)

Cuadro 3. Coberturas y Valores promedio de las Longitudes de Intervalos LVPLA.

	Ξ_1		Ξ_2		Ξ_3	
	cober.	long. prom.	cober.	long. prom.	cober.	long. prom.
μ_1	0.90	(0.3600)	0.89	(0.4500)	0.93	(0.3000)
μ_2	0.97	(0.2400)	0.88	(0.2600)	0.95	(0.2300)
φ_1	0.97	(0.3500)	0.89	(0.3800)	0.98	(0.3200)
φ_2	0.98	(0.2200)	0.92	(0.2200)	0.98	(0.2200)

Cuadro 4. Coberturas y Valores promedio de las Longitudes de Intervalos EM.

5. Federico

Hablar sobre un colega que se ha ido y sobre todo de uno con una trayectoria profesional de magnitud inconmensurable, no resulta sencillo. Forjador y promotor de un área de las matemáticas la cual, sin su presencia, no hubiera sido igual en este país, excelente profesor e investigador, buen padre de familia y amoroso abuelo, celoso defensor del enfoque frequentista, pero además un ser humano con una facilidad natural y constructiva para acercarse a aquellos con los que él trataba diariamente. En la lista de colegas, alumnos y otras personas que tuvimos la fortuna de convivir con Federico, había desde quienes, dependiendo de su problemática humana, recibían consejos, autoanálisis amenizados con chistes y útiles direcciones de médicos, hasta quienes aprovecharon su gran calidad académica para crecer en sus carreras profesionales (me incluyo en todos los rubros). La mayor parte del tiempo, sus chistes iban encaminados a prevenirnos de que podíamos incurrir en una situación difícil: *No te vaya a pasar lo que al del chiste...*, era el preámbulo con el que Federico te preparaba para reír (junto con él), no tanto del chiste en sí, sino de imaginarnos en una posición como la que él narraba, pero en nuestro contexto. Para recibir estas consultas con el doctor O'Reilly, no siempre era necesario tocar a su puerta, a veces, cuando él se daba cuenta de que no estábamos bien o cuando consideraba prudente retomar algún pendiente de nuestras conversaciones, se acercaba. Aunque en ocasiones él usaba sus bromas más escabrosas (celebradas y recordadas por más de uno) con el fin de obtener

nuestra atención, sabía dejar de lado ese estilo para dar consejos en forma que podía llegar a ser poética. Una noche conversábamos en una reunión, cuando notó mi inconformidad con mi forma de relacionarme socialmente y, cuando en el fondo sonaban las notas de una melodía de Clapton que él puso, me dijo: *Las relaciones humanas se manejan como cuando te metes al mar, no entras a la profundidad de súbito, sino que vas dando pasitos para medir si puedes seguir o si debes detenerte. Pero la peor cosa que harías es no meterte al mar.*

Bibliografía

- [1] K. Anaya y F. O'Reilly, «Fisher's observed information matrix for location and scale parameters under type I censoring», *Communications in Statistics, Theory and Methods*, núm. 29, 2000, 1527–1537.
- [2] D. Campos, C. Martínez, A. Contreras-Cristán y F. O'Reilly, «Inference for Mixtures of Distributions for Centrally Censored Data with Partial Identification», *Communications in Statistics, Theory and Methods*, núm. 39, 2010, 2241–2263.
- [3] A. Contreras-Cristán, E. Gutierrez-Peña y F. O'Reilly, «Inference Using Latent Variables for Mixtures of Distributions for Censored Data with Partial Identification», *Communications in Statistics, Theory and Methods*, núm. 32, 2003, 749–774.
- [4] E. Díaz Francés, «Scientific application of maximum likelihood in multiparametric problems», tesis de doctorado, CIMAT, Mexico, 1998.
- [5] B. Efron y D. Hinkley, «Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher Information», *Biometrika*, vol. 65, núm. 3, 1978, 457–487.
- [6] P. McCullagh y R. Tibshirani, «A simple Method for the Adjustment of Profile Likelihoods», *Journal of The Royal Statistical Society*, núm. 52, 1990, 325–344.
- [7] W. Meeker y L. Escobar, *Statistical Methods for Reliability data*, John Wiley & Sons, New York, 1998.
- [8] W. Mendenhall y R. Hader, «Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data», *Biometrika*, núm. 45, 1958, 504–520.
- [9] M. Tanner, *Tools for Statistical Inference*, 3.^a ed., Springer Verlag, New York, 1996.
- [10] D. J. Venzon y S. H. Moolgavkar, «A method for computing profile-likelihood-based confidence intervals», *Applied Statistics*, núm. 37, 1988, 87–94.