

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

**POSGRADO EN CIENCIAS  
MATEMÁTICAS**

**FACULTAD DE CIENCIAS**

**USO DE LA FUNCIÓN DE VEROSIMILITUD EN  
MODELOS DE CENSURA ALEATORIA Y NO  
ALEATORIA**

**T E S I S**

**QUE PARA OBTENER EL GRADO ACADÉMICO DE**

**DOCTOR EN CIENCIAS MATEMÁTICAS**

**P R E S E N T A**

**M. EN C. DANTE GABRIEL CAMPOS SALIDO**

**DR. FEDERICO JORGE O'REILLY TOGNO**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **Agradecimientos:**

Dedico esta tesis a toda la gente que me apoyo para llegar a alcanzar este objetivo, en especial a mi familia por el apoyo y comprensión que me dio en estas circunstancias. Espero que este no sea el último objetivo que supere en la vida.

Quiero agradecer a mi asesor de tesis Dr. Federico O'Reilly, quien tuvo la paciencia de revisar varias veces la tesis y corregir todos mis errores.

También quiero agradecer a la Dra. Silvia Ruiz-Velasco, quien me apoyo siempre que la necesite.

## ÍNDICE:

INTRODUCCIÓN	1
I: “ANTECEDENTES”	5
1.1 CENSURA ALEATORIA	5
1.1.1 CENSURA TIPO I	6
1.1.2 CENSURA TIPO II	10
1.2 CENSURA ALEATORIA	11
1.3 ANÁLISIS DE SUPERVIVENCIA	16
1.3.1 FUNCIÓN DE RIESGO	16
1.3.2 PROPIEDAD DE PÉRDIDA DE MEMORIA	18
1.3.3 ALGUNAS FUNCIONES DE SUPERVIVENCIA CON SU CORRESPONDIENTE FUNCIONES DE RIESGO	20
1.4 ESTIMADOR KAPLAN-MEIER	24
1.4.1 ESTIMADOR PARA LA VARIANZA DE KAPLAN MEIER	26
II: “MANEJO DE DATOS PERTENECIENTES A K POBLACIONES BAJO CENSURA NO ALEATORIA PARCIALMENTE IDENTIFICADOS”	30
2.1 PLANTEAMIENTO DEL PROBLEMA	31
2.2 PROFILE LIKELIHOOD ANALYSIS (PLA)	33
2.3 LATENT VARIABLE PROFILE LIKELIHOOD ANALYSIS (LVPLA)	34
2.4 EJEMPLO CON DOS POBLACIONES	37
2.5 EJEMPLO CON TRES POBLACIONES	44
2.6 SIMULACIONES DE LOS DOS EJEMPLOS ANTERIORES	50
III: “MANEJO DE DATOS PERTENECIENTES A K POBLACIONES BAJO CENSURA ALEATORIA PARCIALMENTE IDENTIFICADOS”	51
3.1. PROFILE LIKELIHOOD ANALYSIS (PLA)	56
3.2. LATENT VARIABLE PROFILE LIKELIHOOD ANALYSIS (LVPLA)	58
3.3 EJEMPLO CON DOS POBLACIONES	62
3.4 EJEMPLO CON TRES POBLACIONES	68
IV: “EJEMPLO”	74
4.1 EJEMPLO CON VALOR EXTREMO	75
4.2. EJEMPLO CON DISTRIBUCIÓN NORMAL	81
4.3 EJEMPLO CON DISTRIBUCIÓN WEIBULL	85

V: “SIMPLIFICACIÓN DEL MÉTODO LVPLA BAJO CENSURA ALEATORIA”	90
5.1. DEMOSTRACIÓN DE LA SEPARACIÓN DE ACUERDO A SU MAGNITUD POR EL MÉTODO LVPLA	90
5.2 PASOS PARA SIMPLIFICAR EL MÉTODO LVPLA	93
5.2.1 PASO 1	94
5.2.2 PASO 2	95
5.3 EJEMPLO CON DOS POBLACIONES	95
5.4 EJEMPLO CON TRES POBLACIONES	103
5.5 EJEMPLO CON DOS POBLACIONES DONDE EXISTE MÁS DE UNA BARRERA	104
5.6 EJEMPLO CON TRES POBLACIONES DONDE EXISTEN CUATRO BARRERAS	108
5.7 ANÁLISIS DE LA ESTABILIDAD DE LA SIMPLIFICACIÓN DEL MÉTODO LVPLA	113
CONCLUSIONES	116
BIBLIOGRAFÍA	118
APENDICE 1: DATOS DEL EJEMPLO DEL CAPÍTULO 4	120
APENDICE 2: PROGRAMAS	121

## **INTRODUCCIÓN**

En la práctica, especialmente en medicina, es común que se desee conocer el tiempo de muerte de un grupo de pacientes bajo una enfermedad terminal, o bien, en la industria, es de interés conocer el tiempo de falla de un conjunto de componentes.

Como suele no ser posible observar durante todo el tiempo, por alguna razón, se decide sólo registrar el tiempo de muerte o falla de un componente solo para un subconjunto de las unidades. Este proceso es el que se conoce como censurar los datos. Lo anterior se puede hacer de varias formas distintas.

La primera forma: Censura Tipo I, consiste en seleccionar un intervalo de tiempo conocido como región de no censura. Para las observaciones que “caen” en dicha región se les registrará el tiempo de muerte o falla, mientras que para las observaciones que “caen” fuera de ésta, es decir, la región de censura, no se registrará sus tiempos de muerte o falla, sino sólo la existencia de dichos valores. No conocemos sus tiempos exactos de falla, pero si sabemos que dichos tiempos pertenecen a la región de censura.

En la segunda forma de censurar los datos se escoge el número de observaciones que van a ser registradas, por lo que solo registraremos algunos tiempos de falla, mientras que de las restantes, solo conoceremos la existencia de dichas observaciones y que su tiempo de muerte o falla es distinto de cualquiera de las fallas registradas, en este caso se sabe de antemano el número de observaciones registradas. A este tipo de censura se le conoce como Censura Tipo II.

Existe un tercer tipo de censura, llamado Censura Aleatoria, donde el investigador no selecciona de antemano el tiempo de censura, como es el caso de la Censura Tipo I, ni el número de datos a observar, como es el caso de la Censura Tipo II, sino que el hecho de que se registre o no el tiempo de muerte o falla de una unidad depende de un conjunto de eventos que se da de manera independiente y aleatoria.

Por ejemplo, supongamos que tenemos un conjunto de pacientes con una enfermedad terminal y queremos estudiar el tiempo de muerte de dichos individuos; sin embargo, puede suceder que los pacientes no continúen en el estudio debido a que mueran a causa de un accidente, o por mudarse de domicilio y quedar inaccesibles. En este caso, no sabemos de antemano, el número de observaciones que se van a registrar, además de que no existe una región fija en la cual se van a registrar o no, los diferentes tiempos de muerte de los pacientes por la enfermedad en bajo estudio.

En ocasiones se tiene que la causa de muerte de un paciente o la falla de un componente, que se quiere analizar, no es única, sino que pueden ser varias las causas de muerte o falla, por lo que se tienen varias poblaciones. Si se registró el tiempo de muerte o falla, además de conocer dicho valor sabemos la causa de muerte o falla, por lo que sabemos a que población pertenece. Si no conocemos el tiempo de muerte o falla, ya que se censuró la observación, sólo conocemos de su existencia, pero no sabemos su valor exacto ni a cual de las poblaciones bajo estudio pertenece.

Este problema ya se ha estudiado, bajo Censura tipo I, en Diaz-Frances E, (1998), donde se utiliza el método “Análisis de Verosimilitud Perfil”, conocido también como PLA, por sus iniciales en inglés. Se modeló con una mezcla o (mixturas) de distribuciones con parámetros de localización y escala para poder obtener los estimadores máximo verosímiles, se utiliza la verosimilitud tal cual y también se encuentra una forma de estimar el número de datos que pertenece a cada población.

En Campos *et al* (2007) se utiliza, también bajo Censura tipo I, el método “Análisis de Verosimilitud Perfil utilizando Variables Latentes”, o LVPLA (Latent Variable Profile Likelihood Analysis), para estimar los parámetros y asignar el número de observaciones censuradas que pertenecen a cada población, a través de la verosimilitud perfil. Además de que se encuentran las ventajas que tiene el método LVPLA sobre el método PLA.

Tanto el método PLA, como el LVPLA toman en cuenta los valores de las observaciones registradas, así como la población a la que pertenecen, además de que consideran los datos censurados. El método PLA utiliza la función de verosimilitud, mientras que en el método LVPLA se construye una especie de verosimilitud perfil, que facilita la construcción de la matriz observada de Fisher.

La diferencia entre estos dos métodos consiste en que el método PLA utiliza más variables: el número de parámetros para cada población, así como la proporción de datos que pertenecen a cada población, por lo que el proceso mediante este método es muy complicado. Por otra parte, el método LVPLA supone que se conoce el número de datos censurados de cada población, lo cual significa que se reduce el número de variables con que se va a trabajar, y la función de verosimilitud se puede descomponer en varias funciones de verosimilitud perfil, las cuales son más sencillas de trabajar. Como, en realidad, no se conoce el número de datos censurados de cada población y es lo que estamos suponiendo, es necesario repetir este método para todas las posibles combinaciones de asignación de los censurados a la población.

El propósito de este trabajo es resolver el siguiente problema: supongamos que tenemos un conjunto de datos los cuales sabemos que pueden pertenecer a una y solamente una de las poblaciones dadas. De estos datos tenemos un subconjunto en el que fue registrado su valor, y además de su valor, conocemos la población a la que pertenece. Mientras que del restante conjunto de datos que fue censurado aleatoriamente, no sabemos ni el valor de la observación ni a que población pertenece. Con esta información, se desarrollará una variante del método LVPLA para estimar el valor de los parámetros, asignar cada uno de los datos censurados a una de las poblaciones bajo cuestión, y por último para realizar inferencias, como por ejemplo, construir intervalos de confianza y resolver pruebas de hipótesis sobre los valores de los parámetros. Por último se comparará con el método PLA y se verán sus ventajas sobre éste.

En el Capítulo 1 “Antecedentes”, analizaremos más a fondo la Censura Tipo I, Censura Tipo II y Censura Aleatoria, además de que se definirá la función de verosimilitud en cada uno de los casos. Veremos lo que se ha hecho anteriormente con los datos que se encuentran bajo cualquier tipo de censura, hablaremos del Análisis de Supervivencia y del Estimador Kaplan-Meier.

En el Capítulo 2 “Manejo de Datos Censurados Pertenecientes a Mezclas”, utilizaremos los métodos PLA y LVPLA para estimar los parámetros de cada una de las poblaciones mediante la asignación de un número de observaciones censuradas a cada una de las poblaciones.

En el Capítulo 3 “Manejo de Datos Censurados Aleatoriamente Pertenecientes a Mezclas” se resolverá el problema anterior, pero con datos bajo censura aleatoria y se analizarán las diferencias que existen con respecto al caso anterior, esto es una extensión del método LVPLA para datos censurados de forma aleatoria y es la principal aportación de la presente tesis.

En el Capítulo 4 “Ejemplos” se llevarán a cabo diversos ejemplos utilizando los resultados del capítulo anterior y se verá que tan efectivo es este método, además de que se construirán intervalos de confianza.

Por último en el Capítulo 5 “Simplificación del Método LVPLA”, se busca hacer una simplificación del método LVPLA, para hacer la búsqueda más eficiente. Lo anterior es una simplificación del método desarrollado en el capítulo 3 y constituye otra de las aportaciones de la tesis.

## **1. ANTECEDENTES**

El propósito de este Capítulo es definir de manera amplia el proceso de censura en los datos ya sea de forma no aleatoria o aleatoria, además de que se analizarán las distintas formas desarrolladas anteriormente, para manejar datos identificados parcialmente. Lo que se pretende hacer es revisar las formas en que se pueden censurar los datos, y los principales resultados que se han utilizado para manejar dichos datos.

Este capítulo se divide en las siguientes Secciones: “Censura No Aleatoria”, donde se definirán las dos formas de censurar datos de forma no aleatoria, las regiones de censura y no censura que se desprenden de dicha definición, la construcción de la función de verosimilitud y la importancia de ésta en lo que se refiere al presente trabajo.

En la Sección “Censura Aleatoria” se hará lo mismo que se hizo en la sección anterior pero con este tipo de censura.

En la Sección “Análisis de Supervivencia” se revisan las funciones de supervivencia y de riesgo, sus características principales y como se construyen a partir de una función de densidad.

Por último, la Sección “Kaplan Meier” es para describir dicho estimador y se verá la utilidad que tiene éste bajo la censura aleatoria.

### **1.1 CENSURA (NO ALEATORIA)**

Supongamos que se desea analizar el tiempo de muerte de un grupo de pacientes que tienen una enfermedad específica, o bien, deseamos ver el tiempo de falla de varias componentes.

Por diversas circunstancias, ya sea por falta de dinero o tiempo, no es posible extender el periodo de observación hasta el momento de muerte del último paciente o el tiempo

de falla de la última componente, por lo que se decide restringir el período de observación. Este proceso es el que se conoce como censurar datos de forma no aleatoria y existen dos formas para ello:

### 1.1.1 CENSURA TIPO I:

En este tipo de censura la persona que está llevando a cabo el experimento decide establecer una región temporal en la cual se van a registrar los tiempo de muerte o falla de los pacientes o componentes respectivamente, dicha región se conoce como región de no censura, mientras que su complemento se define como la región de censura.

Entonces, contamos con dos regiones de importancia, que son ajenas entre sí: la región de no censura, que es el periodo de observación que el investigador seleccionó para registrar el tiempo de falla o muerte de cada unidad observada, y la región de censura, que es donde solo sabremos la existencia del dato, y aunque no sabemos el valor exacto del tiempo de muerte o falla de esa unidad, sí sabemos entre que valores está dicho tiempo de muerte o falla.

De estas dos regiones podemos definir varios tipos de regiones de censura:

- 1) Censura por la derecha. Este tipo de censura es el más común en la práctica porque se da de manera más natural. Se fija un número  $C$ , perteneciente a los reales y si el tiempo de muerte o falla ocurre antes de dicho tiempo  $C$ , entonces se registra el valor de la observación. Por el contrario si ocurre después de un tiempo  $C$ , además de saber la existencia de la observación sabremos que el tiempo de falla es mayor a  $C$ . La región de no censura es:  $(-\infty, C]$ , mientras que la región de censura queda como  $(C, \infty)$ .
- 2) Censura por la izquierda. Este tipo de censura se da, por ejemplo, cuando el investigador no comenzó las observaciones desde el principio sino después de un tiempo  $C$ . En este caso la región de no censura es:  $[C, \infty)$ , mientras que la región de censura queda como:  $(-\infty, C)$ .

- 3) Censura por ambos lados. Este tipo de censura es una combinación de los dos casos anteriores, ya que se decide censurar tanto a la izquierda como a la derecha. Los tiempos de muerte o falla se registran si estos ocurren después de un tiempo  $C_l$ , pero antes de un tiempo  $C_u$ . Por lo que la región de no censura queda como  $[C_l, C_u]$ , mientras que la región de censura consiste en dos intervalos ajenos entre sí:  $(-\infty, C_l) \cup (C_u, \infty)$ . Un ejemplo clásico de este tipo de censura ocurre cuando se desea observar el tiempo en que un grupo de chimpancés bajan de un árbol. Cuando llega el investigador, algunos chimpancés ya bajaron del árbol, estos datos están censurados por la izquierda, por otra parte cuando el investigador se retira podría haber algunos que no han bajado aún, por lo que estarían censurados por la derecha.
- 4) Censura en un intervalo. Este caso es el contrario al anterior. Los datos se registrarán si ocurren antes de un tiempo  $C_l$ , o después de un tiempo  $C_u$ . En este caso la región de censura queda como:  $(C_l, C_u)$ , por lo que la región de no censura es  $(-\infty, C_l] \cup [C_u, \infty)$ . Un ejemplo de este tipo de censura es cuando el investigador tiene que suspender las observaciones en un periodo de tiempo  $(C_l, C_u)$ , por lo que en ese intervalo de tiempo no se sabe el tiempo exacto de muerte o falla, para las observaciones que ocurrieron en dicho intervalo.
- 5) Censura en varios intervalos. Este tipo de censura es una generalización del anterior, y de hecho se puede considerar como generalización de todos los casos anteriormente vistos. En este caso el investigador no pudo registrar los tiempos de muerte o falla, en “s” intervalos de tiempo ajenos entre sí, por lo que únicamente tiene información en el complemento la unión de dichos intervalos. En este caso la región de censura es:  $(C_1^1, C_u^1) \cup (C_1^2, C_u^2) \dots \cup (C_1^s, C_u^s)$ . Mientras que la región de no censura es la intersección de los complementos de dichos intervalos.

FUNCIÓN DE VEROSIMILITUD PARA DATOS BAJO CENSURA TIPO I EN UNA  
DISTRIBUCIÓN PERTENECIENTE A LA FAMILIA DE LOCALIZACIÓN Y  
ESCALA

Vamos a obtener la función de verosimilitud para todos los casos, comenzaremos con el caso de censura por la derecha y se verá que los demás casos saldrán de forma parecida.

Supongamos que tenemos  $x_1, x_2, \dots, x_n$ , datos independientes registrados, además tenemos  $r$  observaciones que fueron censuradas, todas ellas pertenecientes a una misma función de distribución con parámetros de localización y escala. Denotamos a  $f$  como la función de densidad cuyos parámetros son:  $\mu$  y  $\sigma$  que son de localización y escala respectivamente.

Al suponer una densidad perteneciente a una familia de localización y escala tenemos que la densidad es:  $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$ , con  $\mu$  y  $\sigma$ , parámetros de localización y escala respectivamente, y la función de distribución queda como:  $F\left(\frac{x-\mu}{\sigma}\right)$ .

De los datos censurados, en censura por la derecha, solo sabemos que estos son mayores a  $C$ , por lo que la función de verosimilitud queda como:

$$L(\mu, \sigma, x) = \left(1 - F\left(\frac{C - \mu}{\sigma}\right)\right)^r \frac{1}{\sigma^n} \prod_{i=1}^n f\left(\frac{x_i - \mu}{\sigma}\right).$$

Si utilizamos las siguientes transformaciones:

$$z_i = \frac{(x_i - \mu)}{\sigma}, \quad i=1, \dots, n \quad C^* = \frac{(C - \mu)}{\sigma}, \quad (1)$$

Tenemos que la expresión de la función de verosimilitud queda como:

$$L(\mu, \sigma, x) = \frac{1}{\sigma^n} (1 - F_0(C^*))^r \prod_{i=1}^n f_0(z_i)$$

Si la censura es por la izquierda la función de verosimilitud es:

$$L(\mu, \sigma, x) = \frac{1}{\sigma^n} (F_0(C^*))^r \prod_{i=1}^n f_0(z_i).$$

Utilizando, nuevamente, la transformación de (1).

Para la censura por ambos lados tenemos:

$$L(\mu, \sigma, x) = \frac{1}{\sigma^n} (F_0(C_l^*))^{r_1} (1 - F_0(C_u^*))^{r_2} \prod_{i=1}^n f_0(z_i).$$

Con  $r_1$ , el número de observaciones censuradas a la izquierda y  $r_2$  el número de observaciones censuradas a la derecha. Además de que utilizamos las siguientes transformaciones adicionales:

$$C_l^* = \frac{(C_l - \mu)}{\sigma}, \quad C_u^* = \frac{(C_u - \mu)}{\sigma}. \quad (2)$$

Si la censura es en un intervalo la función de verosimilitud es:

$$L(\mu, \sigma, x) = \frac{1}{\sigma^n} (F_0(C_u^*) - F_0(C_l^*))^r \prod_{i=1}^n f_0(z_i).$$

Utilizando las transformaciones adicionales (2) del caso de censura por ambos lados.

Por último, para el caso más general, es decir censura en varios intervalos tenemos:

$$L(\mu, \sigma, x) = \frac{1}{\sigma^n} \prod_{j=1}^s (F_0(C_{u_j}^*) - F_0(C_{l_j}^*))^{r_j} \prod_{i=1}^n f_0(z_i)$$

Con  $r_j$  el número de datos censurados que pertenecen al  $j$ 'ésimo intervalo,  $[C_1^j, C_u^j]$ , el  $j$ 'ésimo intervalo de censura, y  $s$  el número de intervalos de censura, además de las siguientes transformaciones:

$$C_{l_j}^* = \frac{(C_{l_j} - \mu)}{\sigma}, \quad C_{u_j}^* = \frac{(C_{u_j} - \mu)}{\sigma}, \quad \text{con } j=1,2,\dots,s. \quad (3)$$

Como se verá más adelante, la función de verosimilitud será de vital importancia para el manejo de datos censurados.

### 1.1.2 CENSURA TIPO II:

En este caso, en vez de seleccionar el intervalo de tiempo durante el cual registraremos el valor del dato, se escogerá el número de datos de los cuales se registrará su valor.

Supongamos que se decide registrar las  $n$  primeras fallas. Entonces tenemos  $x_1, x_2, \dots, x_n$ , observaciones que suponemos pertenecen a una familia de localización y escala. En este caso es conveniente utilizar estadísticas de orden, es decir,  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ .

Supongamos, además que tenemos  $r$  observaciones no registradas, en este caso sabemos que cualquiera de estas observaciones es mayor a  $x_{(n)}$ .

Si decidiéramos, por alguna causa, registrar las  $n$  últimas fallas, tendríamos que cualquiera de las observaciones no registradas es menor en  $x_{(1)}$ .

FUNCIÓN DE VEROSIMILITUD PARA DATOS BAJO CENSURA TIPO II EN  
UNA DISTRIBUCIÓN PERTENECIENTE A LA FAMILIA DE LOCALIZACIÓN Y  
ESCALA

En caso de que los datos estén censurados por la derecha, utilizando nuevamente la estandarización, tenemos que la función de verosimilitud es:

$$L(\mu, \sigma, x) = (1 - F_0(x_{(n)}, \mu, \sigma))^r \frac{1}{\sigma^n} \prod_{i=1}^n f_0(x_{(i)}, \mu, \sigma).$$

Lo cual se puede escribir de la siguiente forma:

$$L(\mu, \sigma, z) = \frac{1}{\sigma^n} (1 - F_0(z_{(n)}))^r \prod_{i=1}^n f_0(z_{(i)})$$

Utilizando la transformación en (1).

Si la censura es por la izquierda tenemos que la función de verosimilitud queda como:

$$L(\mu, \sigma, x) = (F_0(x_{(1)}, \mu, \sigma))^r \frac{1}{\sigma^n} \prod_{i=1}^n f_0(x_{(i)}, \mu, \sigma), \text{ o bien, si utilizamos la estandarización}$$

$$\text{tenemos } L(\mu, \sigma, z) = \frac{1}{\sigma^n} (F_0(z_{(1)}))^r \prod_{i=1}^n f_0(z_{(i)}).$$

## 1.2 CENSURA ALEATORIA

En este tipo de censura el investigador no escoge de antemano, ni el tiempo de observación de los datos, ni el número de observaciones que se van a registrar, sino que la causa de censura es aleatoria, es decir, que se da por causas ajenas al investigador.

Supongamos que estamos realizando un estudio acerca del tiempo de muerte, de un grupo de pacientes debido a una enfermedad específica. En este caso puede suceder que

haya pacientes que abandonan el estudio o mueren debido a otra causa que no nos interesa. Entonces es imposible conocer de antemano el tiempo de abandono del estudio, este proceso es el que se conoce como censurar datos de forma aleatoria.

Para este tipo de censura necesitamos considerar dos funciones de densidad, la primera que corresponde al tiempo de muerte que denotaremos por  $f_0(x, \mu, \sigma)$ , y suponemos que pertenece a una familia de localización y escala, donde  $\mu$  es el parámetro de localización y  $\sigma$  el parámetro de escala, la función de distribución se denota por  $F_0(x, \mu, \sigma)$ . La segunda función de densidad,  $g$ , está relacionada al de censura, en este caso la función de distribución se denota por  $G$ . Ambas funciones las supondremos conocidas, aunque en la primera los parámetros son desconocidos. La causa de censura la supondremos independiente respecto de la causa de muerte.

Denotamos a  $x_i$ , como el tiempo de muerte de la  $i$ 'ésima observación mientras que  $t_i$  es el tiempo de censura de la  $i$ 'ésima observación. Además de estas dos variables se construye una función indicadora,  $\delta$ , la cual toma los siguientes valores:

1: si el valor de la observación fue registrado.

0: si el valor de la observación fue censurado, es decir, no fue registrado.

Entonces los datos con los que cuenta la persona que realiza la investigación son de la siguiente forma:

$(W_i, \delta_i)$  con  $i=1, \dots, n$

$W_i$  es el tiempo registrado, ya sea por falla o por censura.

En el caso de la censura por la derecha, lo que se observará es  $W_i = \min(x_i, t_i)$ . Si el tiempo de falla ocurrió antes de la censura, sabemos el tiempo exacto de la ocurrencia de la falla, por lo que  $\delta$  vale 1. En cambio, si la causa de censura ocurrió antes del tiempo de falla, no conocemos el valor exacto en el que hubiera ocurrido la falla, sin embargo sabemos que el tiempo de falla sería mayor al tiempo de censura, en este caso  $\delta=0$ .

Sean  $X$  la variable aleatoria que describe los tiempos de falla de la variable de interés y  $T$  la variable aleatoria que describe los tiempos de censura. Para obtener la función de densidad, es necesario, construir, primero la función de distribución.

$$P(W \leq w) = P(\min(X, T) \leq w) = 1 - P(\min(X, T) \geq w)$$

Para que el mínimo entre ambas variables sea mayor a  $w$ , es necesario que cada una de ellas lo sea conjuntamente, es decir:

$$1 - P(\min(X, T) \geq w) = 1 - P(X \geq w, T \geq w)$$

y por independencia podemos escribir la conjunta como el producto de las marginales:

$$1 - P(X \geq w, T \geq w) = 1 - P(X \geq w)P(T \geq w) = 1 - (1 - F(w, \mu, \sigma))(1 - G(w)).$$

Lo cual lo podemos simplificar de la siguiente manera:

$$1 - (1 - F_0(w, \mu, \sigma))(1 - G(w)) = F_0(w, \mu, \sigma) + G(w) - F_0(w, \mu, \sigma)G(w).$$

Si denotamos a  $R(w, \mu, \sigma)$  la función de distribución de los datos observados se tiene:

$$R(w, \mu, \sigma) = F_0(w, \mu, \sigma) + G(w) - F_0(w, \mu, \sigma)G(w)$$

Para obtener la función de densidad, únicamente tenemos que derivar con respecto a  $w$ , es decir:

$$r(w, \mu, \sigma) = f_0(w, \mu, \sigma) + g(w) - f_0(w, \mu, \sigma)G(w) - F_0(w, \mu, \sigma)g(w) = (1 - G(w))f_0(w, \mu, \sigma) + (1 - F_0(w, \mu, \sigma))g(w).$$

En el caso de que la censura aleatoria se de por la izquierda, el tiempo registrado será el máximo entre el tiempo de falla y el tiempo de censura, por lo que las funciones de distribución y densidad quedan como:

$$R(w, \mu, \sigma) = F_0(w, \mu, \sigma)G(w) \text{ y}$$

$$r(w, \mu, \sigma) = f_0(w, \mu, \sigma)G(w) + F_0(w, \mu, \sigma)g(w) .$$

FUNCIÓN DE VEROSIMILITUD PARA DATOS BAJO CENSURA ALEATORIA  
EN UNA DISTRIBUCIÓN PERTENECIENTE A LA FAMILIA DE  
LOCALIZACIÓN Y ESCALA

Como ya se mencionó anteriormente, en este documento es muy importante poder construir la función de verosimilitud, considerando los datos censurados. Esto ya se hizo en la sección anterior donde éstos se encontraban bajo censura tipo I o II. Vamos a hacer algo similar bajo censura aleatoria. Primero supondremos que estamos bajo el caso de censura aleatoria por la derecha.

Para construir la función de verosimilitud es necesario considerar dos situaciones: que el dato haya sido censurado o no:

Si tenemos el par  $(w_i, \delta_i)$ , y  $\delta_i = 1$ , entonces, sabemos que el dato fue registrado, que el tiempo de la causa de censura es mayor a  $w_i$ . La parte de la verosimilitud para este dato queda:

$$(1 - G(w_i))f(w_i, \mu, \sigma) .$$

Por otra parte, si tenemos el valor  $w_j$  y sabemos que la observación fue censurada tenemos que el tiempo de falla fue mayor a  $w_j$ , por lo que la función de verosimilitud para este tipo de observaciones queda como:

$$(1 - F(w_j, \mu, \sigma))g(w_j) .$$

Si tenemos  $n$  observaciones que fueron registradas y  $r$  observaciones censuradas, todas ellas independientes entre sí, tenemos que la función de verosimilitud queda de la siguiente manera:

$$\prod_{i=1}^n (1 - G(w_i)) f(w_i, \mu, \sigma) \prod_{j=1}^r (1 - F(w_j, \mu, \sigma)) g(w_j).$$

Utilizando el hecho de que viene de una familia de localización y escala tenemos:

$$L(\mu, \sigma, w) \propto \frac{1}{\sigma^n} \prod_{i=1}^n (1 - G(w_i)) f\left(\frac{w_i - \mu}{\sigma}\right) \prod_{j=1}^r (1 - F\left(\frac{w_j - \mu}{\sigma}\right)) g(w_j).$$

Como ya conocemos la distribución de la causa de censura y no nos interesa nada con respecto a esta, podemos simplificar la función de verosimilitud de la siguiente forma:

$$L(\mu, \sigma, w) \propto \frac{1}{\sigma^n} \prod_{i=1}^n f\left(\frac{w_i - \mu}{\sigma}\right) \prod_{j=1}^r (1 - F\left(\frac{w_j - \mu}{\sigma}\right)).$$

Si hacemos algo parecido en el caso de censura aleatoria por la izquierda tenemos que la función de verosimilitud queda como:

$$L(\mu, \sigma, w) \propto \frac{1}{\sigma^n} \prod_{i=1}^n f\left(\frac{w_i - \mu}{\sigma}\right) \prod_{j=1}^r (F\left(\frac{w_j - \mu}{\sigma}\right)).$$

La anterior expresión la podemos ver como en el caso de censura no aleatoria, en el caso más general, donde teníamos distintos intervalos de censura ajenos entre sí.

La diferencia, en este caso, es que aquí los intervalos de censura no son ajenos entre sí, de hecho todos son de la forma:  $[w_j, \infty)$ . Incluso puede ocurrir que se registre una observación  $w_k$ , en el intervalo  $[w_j, \infty)$ , que es la región de censura para otra observación, es decir que las regiones de censura no son ajenas con la región de no censura.

### 1.3 ANÁLISIS DE SUPERVIVENCIA

Para el caso de censura no aleatoria, por la derecha, se dice que una unidad ha sobrevivido un tiempo  $t$ , si una vez transcurrido este tiempo, no se ha presentado la falla de dicha unidad. La función de Supervivencia se define como:

$$S(t) = P(X > t) = 1 - F(t).$$

Donde denotamos a  $F(t)$  la función de densidad que es distinta que en la sección anterior.

Supongamos  $M$ , como el tiempo mínimo de supervivencia que sabemos que una unidad tendrá. Generalmente se supone que  $M=0$ . Para la función de supervivencia se cumple que  $S(M)=1$ .

La función de supervivencia heredera de la función de distribución las siguientes propiedades:

- 1)  $S(t)$  es decreciente.
- 2)  $\lim_{t \rightarrow \infty} S(t) = 0$ .

#### 1.3.1 FUNCIÓN DE RIESGO

La función de riesgo (hazard) denotada como  $h(t)$  se define de la siguiente forma:

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{P(X < t + \Delta | X > t)}{\Delta}.$$

Lo que la función de riesgo cuantifica es la probabilidad de que una unidad muera antes del tiempo  $t+\Delta$ , dado que sobrevivió hasta el tiempo  $t$ , es decir que la falla ocurra en el intervalo  $(t, t+\Delta)$ , y como estamos haciendo tender a cero a  $\Delta$ , entonces se podría ver esto como la tasa de muerte instantánea.

Si queremos encontrar una relación entre la función de riesgo y las funciones de densidad y distribución tenemos que hacer lo siguiente:

$$\begin{aligned}
 h(t) &= \lim_{\Delta \rightarrow 0} \frac{P(X < t + \Delta | X > t)}{\Delta} = \lim_{\Delta \rightarrow 0} \frac{P(t < X < t + \Delta)}{\Delta P(X > t)} = \lim_{\Delta \rightarrow 0} \frac{P(X < t + \Delta) - P(X < t)}{\Delta(1 - P(X \leq t))} = \\
 &= \lim_{\Delta \rightarrow 0} \frac{F(t + \Delta) - F(t)}{\Delta(1 - F(t))} = \frac{1}{(1 - F(t))} \lim_{\Delta \rightarrow 0} \frac{F(t + \Delta) - F(t)}{\Delta} = \frac{f(t)}{(1 - F(t))}.
 \end{aligned}$$

Por otra parte, el último lo podemos expresar como una ecuación diferencial:

$$\frac{f(t)}{(1 - F(t))} = -\frac{\partial}{\partial t} \ln(1 - F(t)).$$

Si igualamos esto último a la función riesgo tenemos:

$$h(t) = -\frac{\partial}{\partial t} \ln(1 - F(t)).$$

Integrando desde  $M$ , que es el valor mínimo que puede tomar la función de supervivencia, hasta  $t$  tenemos:

$$\int_M^t h(s) ds = -\ln(1 - F(t))$$

Ahora despejando, de aquí la función de distribución tenemos:

$$F(t) = 1 - e^{-\int_M^t h(s) ds}.$$

Si queremos, ahora expresarlo en términos de la función de densidad tenemos que derivar ambos términos con respecto a  $t$ :

$$f(t) = h(t)e^{-\int_M^t h(s)ds}.$$

La función de riesgo determina totalmente a la función de densidad, por lo tanto, si dos funciones de densidad tienen la misma función de riesgo, entonces estamos hablando de la misma función de densidad.

De las anteriores expresiones podemos concluir las dos siguientes propiedades de la función riesgo:

- 1)  $h(t) \geq 0$
- 2)  $\int_M^{\infty} h(s)ds = \infty$

La propiedad 1) se da debido a que  $f(t) \geq 0$ , por ser una función de densidad, por lo que:

$h(t)e^{-\int_M^t h(s)ds} \geq 0$ , como  $e^{-\int_M^t h(s)ds} \geq 0$ , entonces  $h(t)$  forzosamente debe ser no negativa.

La segunda propiedad se demuestra de la siguiente manera:

$$\int_M^{\infty} h(s)ds = \lim_{t \rightarrow \infty} \int_M^t h(s)ds = \lim_{t \rightarrow \infty} (-\ln(1 - F(t))) = -\ln(1 - \lim_{t \rightarrow \infty} (F(t))) = -\lim_{t \rightarrow 0} (\ln(t)) = \infty$$

Estas son las dos únicas restricciones que tiene la función de riesgo, ya que esta puede ser creciente o decreciente, como se verá posteriormente.

### 1.3.2 PROPIEDAD DE PÉRDIDA DE MEMORIA

Una función de densidad cumple con la propiedad de pérdida de memoria si se tiene lo siguiente:

$$P(X > t + \Delta | X > t) = P(X > \Delta + M), \text{ para toda } t, \Delta \geq M.$$

Donde  $M$ , como se mencionó anteriormente es el valor mínimo que puede tomar la función de supervivencia, y generalmente se supone que  $M=0$ .

Lo anterior nos dice que la probabilidad de supervivencia al tiempo  $t+\Delta$  dado que ya sobrevivió al tiempo  $t$ , es la misma probabilidad de que sobreviva al tiempo  $\Delta + M$ , es decir que en intervalos de tamaño  $\Delta$ , sea cual sea el intervalo donde nos encontramos, se tiene la misma probabilidad de supervivencia.

### TEOREMA 1:

$$P(X > t + \Delta | X > t) = P(X > \Delta + M) \Leftrightarrow h(t) \equiv \lambda$$

Es decir, la única forma de que se cumpla la propiedad de pérdida de memoria es que la función riesgo sea constante.

Demostración:

1) Supongamos que se cumple la propiedad de pérdida de memoria es decir:

$P(X > t + \Delta | X > t) = P(X > \Delta + M)$ , por una parte tenemos:

$$P(X > t + \Delta | X > t) = \frac{P(X > t + \Delta)}{P(X > t)} = \frac{1 - F(t + \Delta)}{1 - F(t)} = \frac{e^{-\int_M^{t+\Delta} h(s) ds}}{e^{-\int_M^t h(s) ds}} = e^{-\int_t^{t+\Delta} h(s) ds}$$

Por otra parte tenemos que:  $P(X > \Delta + M) = e^{-\int_M^{\Delta+M} h(s) ds}$

Igualando ambas expresiones tenemos:

$$\int_t^{t+\Delta} h(s)ds = \int_M^{\Delta+M} h(s)ds, \text{ ahora derivamos ambos lados con respecto a } t, \text{ por el teorema}$$

fundamental del cálculo:

$h(t + \Delta) - h(t) = 0$ , de lo que concluimos:  $h(t + \Delta) = h(t)$ , y como es valido para toda  $t$ ,  $\Delta \geq M$ , entonces  $h(t)$  es una constante.

2) Supongamos ahora que  $h(t) \equiv \lambda$ , de la primera parte de la demostración vimos que:

$$P(X > t + \Delta | X > t) = e^{-\int_t^{t+\Delta} h(s)ds}, \text{ y por ser una función constante tenemos:}$$

$$P(X > t + \Delta | X > t) = e^{-\lambda\Delta}, \text{ por otra parte tenemos:}$$

$$P(X > \Delta + M) = e^{-\int_M^{\Delta+M} h(s)ds} = e^{-\Delta\lambda}, \text{ por lo tanto son iguales.}$$

### 1.3.3 ALGUNAS FUNCIONES DE SUPERVIVENCIA CON SU CORRESPONDIENTE FUNCIÓN DE RIESGO

1) Exponencial:

$$f(x) = \lambda e^{-\lambda(x-M)} \text{ si } x \geq M. \lambda > 0$$

Esta es la función más sencilla que se puede utilizar al realizar el análisis de supervivencia. El parámetro  $\lambda$ , está relacionado con el tiempo medio de fallas.

$$E(X) = \frac{1}{\lambda} + M, V(X) = \frac{1}{\lambda^2}$$

La función de distribución queda como:

$$F(x) = 1 - e^{-\lambda(x-M)}.$$

Si construimos la función de riesgo tenemos que:  $h(x) = \frac{f(x)}{1-F(x)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda$ .

Como ya vimos, la única forma de que la función de riesgo sea constante es que cumpla con la propiedad de pérdida de memoria, además como la función de riesgo determina la función de densidad entonces se tiene que la única densidad que cumple esto es la exponencial.

## 2) Gamma

$$f(x) = \frac{\lambda^\alpha (x-M)^{\alpha-1} e^{-\lambda(x-M)}}{\Gamma(\alpha)} \text{ si } x \geq M, \alpha, \lambda > 0,$$

Donde la función gamma se define como:  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ .

$\alpha$  se define como el parámetro de forma, mientras que  $\lambda$  es el parámetro de escala, además, si  $\alpha$  es entero se tiene que:

$Y_i \sim \exp(\lambda, M)$ , independientes, entonces  $X = \sum_{i=1}^{\alpha} (Y_i - M) + M \sim \text{Gamma}(\alpha, \lambda, M)$ .

$$E(X) = \frac{\alpha}{\lambda} + M, \quad V(X) = \frac{\alpha}{\lambda^2}$$

La función de distribución, en la Gamma, con  $\alpha$  entero queda:

$$F(x) = 1 - e^{-\lambda(x-M)} \sum_{i=0}^{\alpha-1} \frac{(\lambda(x-M))^i}{\Gamma(i+1)},$$

por lo que la función de riesgo es:

$$h(x) = \frac{\lambda^\alpha (x-M)^{\alpha-1} e^{-\lambda(x-M)}}{\Gamma(\alpha)} = \frac{\lambda^\alpha (x-M)^{\alpha-1}}{\Gamma(\alpha)} \cdot \frac{1}{e^{-\lambda(x-M)} \sum_{i=0}^{\alpha-1} \frac{(\lambda(x-M))^i}{\Gamma(i+1)}} = \frac{\lambda^\alpha (x-M)^{\alpha-1}}{\Gamma(\alpha)} \cdot \frac{1}{\sum_{i=0}^{\alpha-1} \frac{(\lambda(x-M))^i}{\Gamma(i+1)}}.$$

En el caso  $\alpha$  entero, la función de riesgo tiene las siguientes propiedades:

- 1)  $h(x) \leq \lambda$ , con  $x \geq M$
- 2)  $\lim_{x \rightarrow \infty} h(x) = \lambda$
- 3)  $h(x)$  es una función creciente.

### 3) Weibull

$$f(x) = \lambda \gamma (x-M)^{\gamma-1} e^{-\lambda(x-M)^\gamma} \quad \text{si } x \geq M, \lambda > 0 \text{ y } \gamma > 0.$$

La densidad Weibull tiene la siguiente relación con la densidad exponencial:

$$\text{Si } Y \sim \exp(\lambda, M) \text{ entonces } X = (Y - M)^{\frac{1}{\gamma}} + M \sim \text{Weibull}(\lambda, \gamma, M)$$

Si  $\gamma=1$  entonces  $X \sim \exp(\lambda, M)$ .

$$E(X) = \frac{\Gamma(\gamma^{-1} + 1)}{\lambda^{\gamma^{-1}}} + M \quad \text{y} \quad V(X) = \frac{\Gamma(2\gamma^{-1} + 1)}{\lambda^{\gamma^{-2}}} - \frac{(\Gamma(\gamma^{-1} + 1))^2}{\lambda^{\gamma^{-2}}}$$

La función de distribución es:  $F(x) = 1 - e^{-\lambda(x-M)^\gamma}$  si  $x \geq M$ , por lo que la función de riesgo queda como:

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\lambda \gamma (x-M)^{\gamma-1} e^{-\lambda(x-M)^\gamma}}{e^{-\lambda(x-M)^\gamma}} = \lambda \gamma (x-M)^{\gamma-1}.$$

Si  $\gamma < 1$ , la función de riesgo es decreciente y su límite al infinito es cero. (denotadas como DFR por sus siglas en inglés: Decreasing Function Risk)

Si  $\gamma = 1$ , la función de riesgo es constante, estamos en el caso exponencial.

Si  $\gamma > 1$ , la función de riesgo es creciente y su límite al infinito es infinito. (denotadas por IFR).

#### 4) Gumbel

$$f(x) = \frac{\lambda}{\sigma} e^{\frac{(x-\mu)}{\sigma}} e^{-\lambda e^{\frac{(x-\mu)}{\sigma}}}, x \in \mathbb{R}, \text{ con } \mu \in \mathbb{R} \text{ y } \lambda, \sigma > 0.$$

La relación que tiene la función Gumbel con respecto a la Exponencial es la siguiente:

Si  $Y \sim \exp(\lambda, M)$  entonces  $X = \sigma \ln(Y - M) + \mu \sim \text{Gumbel}(\lambda, \mu, \sigma)$ .

$$E(X) = \mu - \sigma\gamma - \ln(\lambda) \quad V(X) = \sigma^2 \frac{\pi^2}{6}$$

$\gamma$  es la constante de Euler-Mascheroni cuyo valor es aproximadamente .57721, la cual se construye de la siguiente forma:

$$\gamma = \lim_{n \rightarrow \infty} \left( \sum_{k=1}^n \frac{1}{k} - \ln(n) \right).$$

La función de distribución queda como:  $F(x) = 1 - e^{-\lambda e^{\frac{(x-\mu)}{\sigma}}}$ .

Por lo que la función de riesgo es: 
$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\frac{\lambda}{\sigma} e^{\frac{(x-\mu)}{\sigma}} e^{-\lambda e^{\frac{(x-\mu)}{\sigma}}}}{e^{-\lambda e^{\frac{(x-\mu)}{\sigma}}}} = \frac{\lambda}{\sigma} e^{\frac{(x-\mu)}{\sigma}}.$$

En este caso la función de riesgo es creciente.

### 5) Pareto

$$f(x) = \frac{kA^k}{x^{k+1}} \text{ con } x \geq A, A > 0 \text{ y } k > 0$$

Si  $Y \sim \exp(1, M)$  entonces  $X = Ae^{-(Y-M)/k} \sim \text{Pareto}(k, A)$

$$E(X) = \frac{Ak}{k-1} \text{ si } k > 1, \quad V(X) = \frac{A^2k}{(k-1)(k-2)} \text{ si } k > 2$$

La función de distribución es:  $F(x) = 1 - \frac{A^k}{x^k}$ , por lo que la función riesgo queda como:

$$h(x) = \frac{\frac{kA^k}{x^{k+1}}}{\frac{A^k}{x^k}} = \frac{k}{x}$$

Esta función es decreciente y cuando  $x$  tiende al infinito, la función de riesgo tiende a cero.

## **1.4 ESTIMADOR KAPLAN-MEIER**

El objeto muestral que es equivalente a la función de distribución empírica cuando la censura no es aleatoria, es el estimador Kaplan-Meier que para el caso de censura aleatoria estima la función de supervivencia.

Para este estimador utilizamos los datos que no fueron censurados y la información que tenemos sobre los datos censurados.

El procedimiento para construir el estimador Kaplan-Meier es el siguiente: suponemos que tenemos un conjunto de unidades bajo observación y definimos  $N$  tiempos de observación  $t_1, t_2, \dots, t_N$ , tales que:

$$0 = t_1 \leq t_2 \leq \dots \leq t_N.$$

Sea  $n_i$  el número de observaciones bajo estudio al tiempo  $t_i$ . Definimos  $d_i$  y  $c_i$  como el número de muertes y unidades censuradas en el intervalo  $[t_i, t_{i+1}]$  respectivamente. De esto podemos ver la siguiente relación  $n_{i+1} = n_i - d_i - c_i$ .

En la sección anterior, expresamos a la función de supervivencia,  $S(t)$ , de la siguiente manera:

$$S(t) = e^{-\int_0^t h(u) du}.$$

Donde  $h(t)$  es la función de riesgo, y como vimos se define como la tasa instantánea de muerte, es decir, la probabilidad de que una unidad muera en el intervalo  $(t, t+\Delta)$ , dado que sobrevivió hasta  $t$ , se divide entre  $\Delta$ , y se hace  $\Delta$  tender a cero.

Si las diferencias que existen entre cada uno de los tiempos es lo suficientemente pequeña, podemos aproximar la integral, con sumas de Riemman:

$$\int_0^t h(u) du \approx \sum_{i=1}^j h(t_i)(t_{i+1} - t_i). \text{ Con } j = \max(i | t_i \leq t).$$

Si sustituimos esta igualdad en la función de supervivencia tenemos que:

$$S(t) = e^{-\int_0^t h(u) du} \approx e^{-\sum_{i=1}^j h(t_i)(t_{i+1} - t_i)} = \prod_{i=1}^j e^{-h(t_i)(t_{i+1} - t_i)}$$

Ahora, utilizamos la definición de la función riesgo:  $h(t_i)$  es la probabilidad de que una unidad muera en  $(t_i, t_i+\Delta)$ , dado que sobrevivió hasta  $t_i$ , haciendo  $\Delta$  tender a cero. Si el intervalo  $(t_i, t_{i+1})$ , es lo suficientemente pequeño, entonces, podemos definir en este caso

a  $h(t_i)$  la probabilidad de que la unidad muera en  $(t_i, t_{i+1})$ , dado que sobrevivió más de  $t_i$ , y dividimos entre la longitud del intervalo.

La probabilidad de que muera en el intervalo en  $(t_i, t_{i+1})$ , dado que sobrevivió más de  $t_i$ ,

es:  $\frac{d_i}{n_i}$ , por lo que  $h(t_i) = \frac{d_i}{n_i(t_{i+1} - t_i)}$ . Si sustituimos tenemos que:

$$\prod_{i=1}^j e^{-h(t_i)(t_{i+1}-t_i)} = \prod_{i=1}^j e^{-\frac{d_i}{n_i(t_{i+1}-t_i)}(t_{i+1}-t_i)} = \prod_{i=1}^j e^{-\frac{d_i}{n_i}}.$$

Si utilizamos la aproximación de Taylor hasta el término lineal tenemos:

$$\prod_{i=1}^j e^{-\frac{d_i}{n_i}} \approx \prod_{i=1}^j \left(1 - \frac{d_i}{n_i}\right) = \prod_{i=1}^j \frac{n_i - d_i}{n_i}$$

Si esto lo expresamos en términos de las  $t_i$ 's tenemos que el estimador Kaplan-Meier queda de la siguiente manera:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$$

En caso de que no existan datos censurados tenemos que:  $n_{i+1} = n_i - d_i$ . El estimador Kaplan Meier queda de la siguiente forma:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_{i+1}}{n_i} = \frac{n_2 n_3 \cdots n_{t+1}}{n_1 n_2 \cdots n_t} = \frac{n_{t+1}}{n_1}$$

La anterior expresión se puede ver como el número de observaciones que sobrevivieron al tiempo  $t$ , entre el número de observaciones totales, es decir que tenemos un estimador de la probabilidad de supervivencia en el caso de censura no aleatoria.

### 1.4.1 ESTIMADOR DE LA VARIANZA PARA KAPLAN-MEIER

Para estimar la varianza del Kaplan-Meier se utilizaron resultados propuestos en Greenwood (1926).

$$\hat{V}(\hat{S}(t)) \approx \hat{S}^2(t) \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}.$$

Demostración:

El estimador Kaplan Meier es:  $\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$ , sea  $\hat{p}_i = \frac{n_i - d_i}{n_i}$ , por lo que el

estimador queda como:  $\hat{S}(t) = \prod_{t_i \leq t} \hat{p}_i$ . Si tomamos logaritmos de ambos lados tenemos:

$$\ln(\hat{S}(t)) = \sum_{t_i \leq t} \ln(\hat{p}_i).$$

Sabemos que  $n_i \hat{p}_i$  se distribuye como Binomial( $n_i, p_i$ ), por lo que:

$Var(\hat{p}_i) = \frac{p_i(1-p_i)}{n_i}$ , como no conocemos el verdadero valor de  $p_i$ , entonces

utilizamos  $\hat{p}_i$ :  $\hat{V}(\hat{p}_i) = \frac{\hat{p}_i(1-\hat{p}_i)}{n_i}$ .

Utilizando la siguiente aproximación de Taylor:  $Var(g(X)) \approx \left(\frac{\partial g}{\partial X}\right)^2 Var(X)$ .

En este caso  $g(X)=\ln(X)$ , por lo que:  $\left(\frac{\partial g}{\partial X}\right)^2 = \frac{1}{X^2}$ .

Valuando en  $X = \hat{p}_i$ , tenemos que:  $Var(\ln(\hat{p}_i)) \approx \left(\frac{1}{\hat{p}_i}\right) \frac{\hat{p}_i(1-\hat{p}_i)}{n_i} = \frac{(1-\hat{p}_i)}{\hat{p}_i n_i}$ .

Si sustituimos  $\hat{p}_i = \frac{n_i - d_i}{n_i}$ , en la anterior ecuación tenemos:

$$Var(\ln(\hat{p}_i)) \approx \frac{(1-\hat{p}_i)}{\hat{p}_i n_i} = \frac{\frac{d_i}{n_i}}{n_i - d_i} = \frac{d_i}{n_i(n_i - d_i)}, \text{ por lo que: } \hat{V}(\ln(\hat{S}(t))) \approx \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}.$$

Ya tenemos la varianza del logaritmo del estimador Kaplan-Meier, para encontrar la varianza del estimador Kaplan-Meier, utilizamos nuevamente la aproximación de Taylor.

$$Var(g(X)) \approx \left(\frac{\partial g}{\partial X}\right)^2 Var(X).$$

En este caso  $g(X) = e^X$ , por lo que  $\left(\frac{\partial g}{\partial X}\right)^2 = e^{2X}$ , si sustituimos  $X = \ln(\hat{S}(t))$ , tenemos:

$$\left(\frac{\partial g}{\partial X}\right)^2 = (\hat{S}(t))^2, \text{ que al sustituir queda como:}$$

$$\hat{V}(\hat{S}(t)) \approx \hat{S}^2(t) \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}.$$

El estimador Kaplan Meier se distribuye asintóticamente como normal cuando el tamaño de muestra es suficientemente grande, por lo que podemos escribir los intervalos de confianza al  $(1-\alpha)\%$  de la siguiente forma:

$$[\hat{S}(t) - z^{1-\alpha/2} \hat{V}(\hat{S}(t)), \hat{S}(t) + z^{1-\alpha/2} \hat{V}(\hat{S}(t))].$$

Muchas veces, cuando no tenemos el tamaño de muestra suficientemente grande, el estimador Kaplan Meier dista de tener una buena aproximación a la distribución normal, pero podemos utilizar una transformación que sí se aproxime a la distribución normal, entonces se aplica dicha transformación para encontrar un intervalo de confianza al nivel deseado.

## **2. MANEJO DE DATOS PERTENECIENTES A K POBLACIONES**

### **BAJO CENSURA (NO ALEATORIA) PARCIALMENTE**

#### **IDENTIFICADOS**

En este capítulo se pretende explicar los métodos Profile Likelihood Analysis (PLA) y Latent Variable Profile Likelihood Analysis (LVPLA), así como su utilidad en resolver problemas de mezclas de poblaciones parcialmente identificados bajo censura no aleatoria.

Supongamos que tenemos  $N$  individuos de los cuales queremos estudiar el tiempo de muerte, de cada uno de ellos. Pero, por alguna razón, no podemos observarlo, sino sólo en un intervalo, al que definimos como región de no censura, mientras que su complemento es la región de censura, este proceso se le conoce como censura tipo I. Tenemos, además, “ $k$ ” poblaciones previamente definidas, que son exhaustivas (todos los individuos pertenecen a alguna población) y ajenas entre sí (un individuo no puede estar en más de una población). Supongamos que el tiempo de falla de cada individuo tiene una distribución perteneciente a un miembro de la familia de localización y escala que depende de la población a la que pertenece.

Además supongamos que cada los datos de cada población son una muestra aleatoria: variables aleatorias independientes e idénticamente distribuidas, lo cual es equivalente a un muestreo aleatorio simple, donde el tamaño de la población es infinito.

Contamos, entonces, con dos situaciones:

- 1) Si la observación ocurrió en la región de no censura, entonces además de registrar su valor, identificamos la población a la que pertenece.
- 2) Si la observación ocurre en la región de censura, solo registramos la existencia del dato, pero, adicionalmente, sabemos que el tiempo de falla pertenece a la región de censura.

Como definimos en el capítulo 1 “Antecedentes”, la región de censura puede ser un intervalo acotado (censura en un intervalo), un intervalo no acotado (censura por la

derecha o por la izquierda) o un conjunto de intervalos (censura por intervalos o censura en varios intervalos).

El propósito de este capítulo es estimar los parámetros de localización y escala de cada una de las “ $k$ ” poblaciones, mediante una asignación de los datos censurados en las “ $k$ ” poblaciones. Para hacer esto, definiremos los métodos *Profile Likelihood Analysis* (PLA) y *Latent Variable Profile Likelihood Analysis* (LVPLA), y analizaremos lo que proponen.

## 2.1 PLANTEAMIENTO DEL PROBLEMA

Sean  $f_i(x, \mu_i, \sigma_i)$  y  $F_i(x, \mu_i, \sigma_i)$  las funciones de densidad y la distribución de la  $i$ 'ésima población, cada una con parámetros de localización  $\mu_i$ , y escala  $\sigma_i$ , y sea  $\pi_i$ , la proporción total de datos que pertenecen a la  $i$ 'ésima población. Estos valores tienen que cumplir con lo siguiente:

- 1)  $\pi_i \geq 0, i=1, \dots, k.$
- 2)  $\sum_{j=1}^k \pi_i = 1.$

Tenemos  $N$  datos que pertenecen a una y sola una de las “ $k$ ” poblaciones. Definimos a  $x_{ij}$ , como la  $j$ 'ésima observación perteneciente a la  $i$ 'ésima población, supongamos que el número total de datos de la población  $i$  es  $n_i$ , y que el número de datos de todas las poblaciones es  $n$ , es decir que la suma de las  $n_i$ 's es  $n$ . Además de estos  $n$  datos identificados tenemos  $r$  datos censurados, de los cuales no sabemos su verdadero valor, ni la población a la que pertenece, solo su existencia, entonces  $r=N-n$ .

La función de verosimilitud queda como:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x}) = (R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}))^r \prod_{i=1}^k \prod_{j=1}^{n_i} f_i(x_{ij}, \mu_i, \sigma_i)$$

Donde:

$$\boldsymbol{\mu} = (\mu_1 \quad \mu_2 \quad \dots \quad \mu_k), \boldsymbol{\sigma} = (\sigma_1 \quad \sigma_2 \quad \dots \quad \sigma_k), \text{ y } \boldsymbol{\pi} = (\pi_1 \quad \pi_2 \quad \dots \quad \pi_k),$$

$R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi})$  es la probabilidad de que un dato haya sido censurado, sea cual fuere la población a la que pertenece.

Por ejemplo, si los datos se encuentran censurados por la derecha:

$$R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}) = 1 - \sum_{i=1}^k \pi_i F_i(C, \mu_i, \sigma_i), \text{ donde } C \text{ es el valor a partir del cual se censura.}$$

Si la censura es por la izquierda tenemos que:  $R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}) = \sum_{i=1}^k \pi_i F_i(C, \mu_i, \sigma_i)$ , en este caso tenemos como  $C$  el valor a partir del cual vamos a registrar las observaciones.

Para el caso en que la censura se dé en un intervalo acotado de la forma  $[C_l, C_u]$

$$\text{entonces } R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}) = \sum_{i=1}^k \pi_i F_i(C_u, \mu_i, \sigma_i) - \sum_{i=1}^k \pi_i F_i(C_l, \mu_i, \sigma_i).$$

La función de distribución  $\sum_{i=1}^k \pi_i F_i(C, \mu_i, \sigma_i)$  es lo que se conoce como una función de distribución mezcla (o mixtura).

Si la región de censura consiste en un conjunto de intervalos ajenos, entonces deberemos saber, cuántos de los datos censurados se encuentra en cada uno de los intervalos ajenos, por lo que la expresión de  $R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi})$ , es similar al caso anterior, solo que considerando los diferentes intervalos de censura correspondientes.

Lo que queremos hacer es utilizar la verosimilitud para poder estimar los parámetros de localización y escala para cada población. Para esto se utilizarán dos métodos, el primero de ellos es el método (PLA)

## 2.2 PROFILE LIKELIHOOD ANALYSIS (PLA)

Este método lo propuso anteriormente Díaz Francés (1998), donde se analizó un ejemplo particular (Mendelhall y Hader 1958), suponiendo dos distribuciones Weibull distintas y se utilizó la censura por la derecha.

Lo primero que se hace en este método es calcular la función de verosimilitud:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x}) \propto (R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}))^r \prod_{i=1}^k \prod_{j=1}^{n_i} \pi_i f_i(x_{ij}, \mu_i, \sigma_i).$$

Ahora, encontramos la log-verosimilitud:

$$l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x}) = (r \ln(R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}))) + \sum_{i=1}^k \sum_{j=1}^{n_i} \ln(\pi_i f_i(x_{ij}, \mu_i, \sigma_i)).$$

Obtenemos las derivadas parciales de la función de log-verosimilitud con respecto a todos los parámetros de interés:  $k$  parámetros de localización,  $k$  parámetros de escala y  $k-1$  parámetros de proporción, ya que la última proporción es redundante. Esto último se hace para encontrar los estimadores máximos verosímiles de dichos parámetros. Dichas expresiones quedan de la siguiente manera:

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x})}{\partial \mu_i} = \frac{r}{R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi})} \frac{\partial R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi})}{\partial \mu_i} + \sum_{i=1}^k \sum_{j=1}^{n_j} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \mu_i}, \quad i=1, 2, \dots, k$$

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x})}{\partial \sigma_i} = \frac{r}{R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi})} \frac{\partial R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi})}{\partial \sigma_i} + \sum_{i=1}^k \sum_{j=1}^{n_j} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \sigma_i}, \quad i=1, 2, \dots, k$$

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x})}{\partial \pi_i} = \frac{r}{R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi})} \frac{\partial R(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi})}{\partial \pi_i} + \sum_{i=1}^k \frac{n_i}{\pi_i}, \quad i=1, 2, \dots, k-1.$$

Una vez encontrados los estimadores máximos verosímiles, se encuentra la matriz observada de Fisher. Tenemos  $k$  parámetros de localización,  $k$  parámetros de escala, y  $k-1$  proporciones estimadas, por lo que la matriz observada de Fisher es de orden  $3k-1$ .

El primer problema que presenta el método es que hay que resolver un sistema de ecuaciones no lineal de tamaño  $3k-1$ , lo cual es muy complicado. El segundo problema que surge es que la matriz observada de Fisher es de orden  $3k-1$ , lo que significa que con un número moderado de poblaciones el orden de dicha matriz es muy grande.

Debido a lo anterior, se propuso otro método, *Latent Variable Profile Likelihood Analysis* (LVPLA).

### **2.3 LATENT VARIABLE PROFILE LIKELIHOOD ANALYSIS (LVPLA)**

Este método se utilizó en Contreras *et al* (2003), donde se supuso una distribución Weibull y censura por la derecha y en Campos *et al* (2007), donde las distribuciones fueron Normales y la censura fue en un intervalo.

Este método consiste en suponer que sabemos el número de datos censurados que corresponde a cada población. Al suponer lo anterior, se eliminan como parámetros a estimar las proporciones, pero lo más importante es que la verosimilitud puede estudiarse por separado descomponiéndola en  $k$  verosimilitudes de 2 parámetros cada una, por lo que la función de verosimilitud queda de la siguiente forma:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{X}) \propto \frac{r!}{r_1! r_2! \dots r_k!} \prod_{i=1}^k (R_i(\mu_i, \sigma_i))^{r_i} \prod_{j=1}^{n_i} f_i(x_{ij}, \mu_i, \sigma_i).$$

En la anterior expresión,  $r_i$  denota el número de datos censurados que pertenecen a la  $i$ 'ésima población (la suposición que se hace en este método).

La ventaja de hacer esto es que la forma de la verosimilitud queda más sencilla, además que podemos trabajar los datos de cada población de manera independiente, Si tomamos logaritmos tenemos:

$$l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{x}) = \sum_{i=1}^k (r_i \ln(R_i(x_i, \mu_i, \sigma_i))) + \sum_{j=1}^{n_j} \ln(f_i(x_{ij}, \mu_i, \sigma_i)) = \sum_{i=1}^k l_i(x, \mu_i, \sigma_i, \pi_i)$$

$$\text{Siendo } l_i(\mu_i, \sigma_i, x) = r_i \ln(R_i(x_i, \mu_i, \sigma_i)) + \sum_{j=1}^{n_j} \ln(f_i(x_{ij}, \mu_i, \sigma_i)).$$

Dado que la derivada es una función lineal podemos trabajar cada una de las log-likelihoods de forma separada, por lo que podemos encontrar los estimadores máximos verosímiles de manera independiente.

De hecho las expresiones de las derivadas quedan de la siguiente forma:

$$\frac{\partial l(\mu, \sigma, x)}{\partial \mu_i} = \frac{r_i}{R_i(\mu_i, \sigma_i)} \frac{\partial R_i(\mu_i, \sigma_i)}{\partial \mu_i} + \sum_{j=1}^{n_j} \frac{1}{f(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \mu_i},$$

$$\frac{\partial l(\mu, \sigma, x)}{\partial \sigma_i} = \frac{r_i}{R_i(\mu_i, \sigma_i)} \frac{\partial R_i(\mu_i, \sigma_i)}{\partial \sigma_i} + \sum_{j=1}^{n_j} \frac{1}{f(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \sigma_i}$$

Como estamos suponiendo que conocemos el número de datos censurados correspondientes a cada población, lo cual no es cierto, es necesario que el vector  $(r_1, r_2, \dots, r_k)$ , se le asignen todas las combinaciones posibles que puede tomar, es decir:

$$r_i \geq 0, i=1, 2, \dots, k$$

$$\sum_{i=1}^k r_i = r$$

El número de posibilidades es:

$$\binom{r+k-1}{r}$$

Para cada una de las combinaciones consideradas obtenemos los estimadores máximo verosímiles de todos los parámetros de localización y escala. Valuamos la función de verosimilitud para cada uno de esos casos, y nos quedamos con el vector  $(r_1^* \ r_2^* \ \dots \ r_{k-1}^*)$  que maximiza la función de verosimilitud total. Equivale lo anterior a suponer que el vector  $(r_1 \ r_2 \ \dots \ r_{k-1})$  es de parámetros (ya no de variables latentes) y que hemos construido su verosimilitud perfil, la cual se maximizó en  $(r_1^* \ r_2^* \ \dots \ r_{k-1}^*)$ .

Una vez encontrado el vector que “estima” el número de datos censurados que le corresponden a cada una de las poblaciones, y sus correspondientes estimadores máximos verosímiles de los parámetros de localización y escala, procedemos a encontrar las matrices de Fisher, para cada una de ellas. Con esto podemos hacer inferencias como lo son la construcción de intervalos de confianza o pruebas de hipótesis utilizando la correspondiente aproximación Normal, pero ahora, partiendo de matrices de dimensión dos.

Las ventajas que presenta este método es que los sistemas de ecuaciones que hay que resolver son de tamaño 2, y en general, las expresiones quedan más sencillas, aunque hay que resolver  $k$  sistemas de ecuaciones. Como las expresiones de las  $k$  matrices de Fisher son más sencillas no se presenta el problema numérico de obtener la inversa del Hessiano.

En caso de que la región de censura no sea un solo intervalo, sino la unión de un conjunto de estos será necesario considerar variables adicionales que nos indican para cada una de las poblaciones como se distribuye el número de datos censurados asignados en cada uno de los “s” intervalos de censura.

El método PLA estima de manera directa la proporción de datos que pertenece a cada una de las poblaciones, para estimar la proporción por el método LVPLA, lo hacemos a través de la estimación del número de datos censurados de cada población, es decir,  $r_i$ .

$$\hat{\pi}_i = \frac{n_i + r_i}{N}.$$

Si quisiéramos realizar inferencias para las proporciones utilizamos una mezcla de Betas, con  $Beta(t, n-t+1)$  y  $Beta(t+1, n-t)$ , donde  $n$  es  $N$  y  $t=n_i+r_i$ . Esta mezcla se propuso en O'Reilly (2003) "Significance Distributions", en el cual se construye dicha distribución con la ponderación de ambas Betas, muy cercano por ciento a la distribución final:  $Beta(t+.5, n-t+.5)$ , usada en un análisis Bayesiano de referencia.

## 2.4 EJEMPLO CON DOS POBLACIONES

Se generan 2 poblaciones normales con 25 datos cada uno, la primera población es una Normal (0,1), mientras que la segunda población es una Normal (1,4).

N(0,1)	N(1,4)
-1.4951	0.08292
-1.2964	2.2514
-1.9277	2.2845
0.2677	-1.6793
2.4305	-0.72868
-0.5491	-1.06140
-1.8519	2.4927
-0.3958	-2.6730
1.6509	1.3257
0.0887	0.86028
1.8898	0.54028
-0.14165	1.63014
-1.0515	0.46025
0.8274	3.2877
0.52467	0.85568
0.1282	3.00851
-0.1340	3.6433
-0.0004	1.8886
-0.3490	1.1387
-0.07010	4.0133
-0.7049	-3.3314
0.2090	2.0395
-0.2842	3.4115
0.8100	-2.1111
0.4620	1.0073

Decidimos tomar como región de censura el intervalo  $[-2,1]$ , por lo que nos quedamos con el siguiente conjunto de datos:

N(0,1)	N(1,4)
-1.4951	0.0830
-1.2964	2.2514
-1.9277	2.2845
2.4305	-1.6793
-0.5491	-0.7287
-1.8520	-1.0614
-0.3958	2.4927
1.6509	-2.6730
0.0887	1.3257
1.8898	1.6301
-0.14167	3.2877
-1.0515	3.0085
0.1282	3.6433
-0.1340	1.8886
-0.0004	1.1387
-0.3490	4.0133
-0.07010	-3.3314
-0.7049	2.0395
-0.2843	3.4115
	-2.1111
	1.0073

Entonces tenemos la siguiente información: 19 datos que pertenecen a la población 1, 21 datos que pertenecen a la población 2, 10 datos que fueron censurados que no sabemos a cual población pertenecen, además de que contamos con la región de censura que es  $[-2,1]$ . Queremos estimar los parámetros de la media y varianza para las dos poblaciones, es decir:  $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$ .

Antes de construir la función de verosimilitud perfil, hacemos la siguientes reparametrizaciones para las varianzas:

$$\sigma_1 = e^{\theta_1} \text{ y } \sigma_2 = e^{\theta_2}.$$

Si lo resolvemos por el método PLA, tenemos que encontrar los parámetros que maximizan la siguiente función de verosimilitud:

$$L(\mu, \varphi, x) \propto \prod_{i=1}^2 \pi_i (\Phi((1 - \mu_i)e^{-\varphi_i}) - \Phi((.2 - \mu_i)e^{-\varphi_i})) \prod_{i=1}^2 e^{-n_i \varphi_i} e^{-\frac{1}{2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i) e^{-2\varphi_i}}$$

Donde  $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ .

Los estimadores máximos verosímiles son:

$$\hat{\mu}_1 = -0.0215, \hat{\varphi}_1 = 0.05171, \hat{\mu}_2 = 0.9762, \hat{\varphi}_2 = 0.6812, \hat{\pi}_1 = 0.5035.$$

Si aplicamos la transformación  $\sigma_i^2 = e^{2\varphi_i}$ , con  $i=1,2$  obtenemos los estimadores máximos verosímiles para las varianzas que son:

$$\hat{\sigma}_1^2 = 1.1089, \hat{\sigma}_2^2 = 3.9055.$$

La matriz calculada de Fisher queda de la siguiente forma:

$$I = \begin{pmatrix} .0063 & .0015 & -.0008 & .0009 & .00124 \\ .0015 & .04633 & -.0005 & .0007 & .0009 \\ -.0008 & -.0005 & .02054 & -.00061 & -.00078 \\ .0009 & .0007 & -.00061 & .1584 & .00089 \\ .00124 & .0009 & -.00078 & .00089 & .02132 \end{pmatrix}$$

Los intervalos de confianza al 95% quedan como:

$$\pi_1 : [0.3479, 0.6590], \quad \mu_1 : [-0.44338, 0.40038], \quad \varphi_1 : [-0.22919, 0.33261], \\ \mu_2 : [0.19613, 0.975627], \quad \varphi_2 : [0.39501, 0.96739].$$

Todos estos intervalos contienen el verdadero valor del parámetro.

El principal problema de este método fue calcular la matriz de Fisher, que en este caso para dos poblaciones queda de tamaño 5 x 5. A continuación lo haremos por el método LVPLA.

Definimos, ahora,  $r_1$  como el número de datos censurados que pertenecen a la población 1, como sabemos que existen 10 datos censurados en total, tenemos, entonces que  $r_2=10-r_1$ . Por lo que las verosimilitudes perfil quedan de la siguiente manera:

$$L_i(\mu_i, \varphi_i, x) = e^{-n_i \varphi_i} e^{-\sum_{j=1}^{n_i} \frac{(x_{ij} - \mu_i)^2 e^{-2\varphi_i}}{2}} (\Phi((1 - \mu_i)e^{-\varphi_i}) - \Phi((.2 - \mu_i)e^{-\varphi_i}))^{r_i} \quad i=1,2.$$

Si obtenemos la log-verosimilitud, tomando logaritmos tenemos:

$$l_i(\mu_i, \varphi_i, x) = -n_i \varphi_i - \frac{1}{2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 e^{-2\varphi_i} + r_i \ln(\Phi((1 - \mu_i)e^{-\varphi_i}) - \Phi((.2 - \mu_i)e^{-\varphi_i}))$$

$i=1,2.$

Cada una de estas verosimilitudes solo dependen de sus correspondientes parámetros de localización y escala, por lo que las derivadas parciales quedan de la siguiente manera:

$$\frac{\partial l_i(\mu_i, \varphi_i, x)}{\partial \mu_i} = \sum_{j=1}^{n_i} (x_{ij} - \mu_i) e^{-2\varphi_i} - r_i e^{-\varphi_i} \frac{(\phi((1 - \mu_i)e^{-\varphi_i}) - \phi((.2 - \mu_i)e^{-\varphi_i}))}{(\Phi((1 - \mu_i)e^{-\varphi_i}) - \Phi((.2 - \mu_i)e^{-\varphi_i}))} \quad i=1,2$$

$$\frac{\partial l_i(x, \mu_i, \varphi_i, x)}{\partial \varphi_i} = -n_i - \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 e^{-2\varphi_i} - r_i e^{-\varphi_i} \frac{((1 - \mu_i)\phi((1 - \mu_i)e^{-\varphi_i}) - \phi((.2 - \mu_i)(.2 - \mu_i)e^{-\varphi_i}))}{(\Phi((1 - \mu_i)e^{-\varphi_i}) - \Phi((.2 - \mu_i)e^{-\varphi_i}))}$$

$i=1,2.$

Si queremos encontrar los estimadores máximos verosímiles tenemos que igualar a cero, lo que equivale a resolver dos sistemas de dos ecuaciones con dos incógnitas, lo cual es más fácil que en el método PLA, donde tendría que resolverse un sistema de ecuación con 5 incógnitas y 5 variables, lo cual es más complicado.

Una vez resuelto ambos sistemas tenemos los estimadores máximo verosímiles para una  $r_1$  fija, el problema es que estamos suponiendo que conocemos el valor de  $r_1$ , lo cual no

es así, por lo que tenemos que hacer este proceso tomando en cuenta todos los valores posibles de  $r_1$ , y por consiguiente de  $r_2$ , y después valuar en la verosimilitud para ver el máximo valor que toma. Al hacer esto tenemos:

$r_1$	$\hat{\mu}_1$	$\hat{\varphi}_1$	$\hat{\mu}_2$	$\hat{\varphi}_2$	L
0	-0.2139	0.1317	0.9023	0.5753	2.232E-25
1	-0.1748	0.1187	0.9122	0.5910	2.7942E-24
2	-0.1393	0.1056	0.9228	0.6072	1.6463E-23
3	-0.1071	0.0925	0.9342	0.624	6.0024E-23
4	-0.0776	0.0795	0.9464	0.6412	1.4982E-22
5	-0.0505	0.0666	0.9595	0.6592	2.6725E-22
6	-0.0256	0.0539	0.9737	0.6779	3.4484E-22
7	-0.0025	0.0414	0.9890	0.6973	3.1767E-22
8	0.0188	0.0290	1.0057	0.7175	1.9991E-22
9	0.03864	0.0169	1.0239	0.7386	7.76E-23
10	0.0571	0.0050	1.0439	0.7606	1.4111E-23

Como se ve en el anterior cuadro, el valor máximo que alcanza la función de verosimilitud es cuando  $r_1=6$ , y los estimadores correspondientes son:

$$\hat{\mu}_1 = -.02557732, \hat{\varphi}_1 = .05390019, \hat{\mu}_2 = .97365682, \hat{\varphi}_2 = .67788041.$$

Si quisiéramos encontrar las varianzas aplicamos las transformaciones correspondientes y tenemos:

$$\hat{\sigma}_1^2 = 1.11382537649468 \text{ y } \hat{\sigma}_2^2 = 3.87971157201281.$$

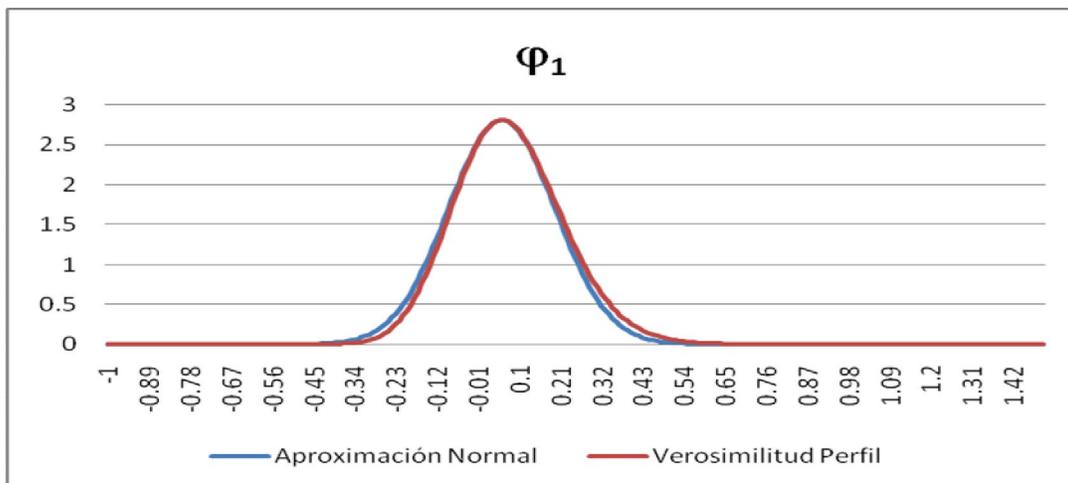
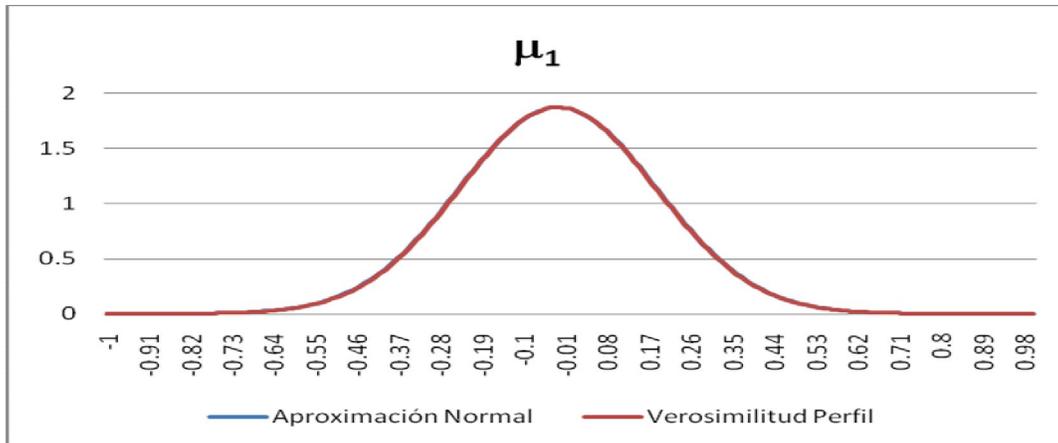
Las matrices observadas de Fisher para cada una de las verosimilitudes son:

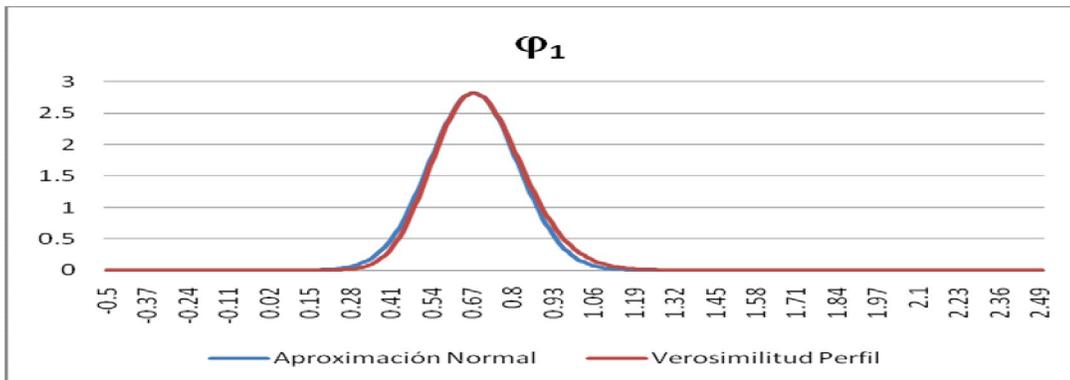
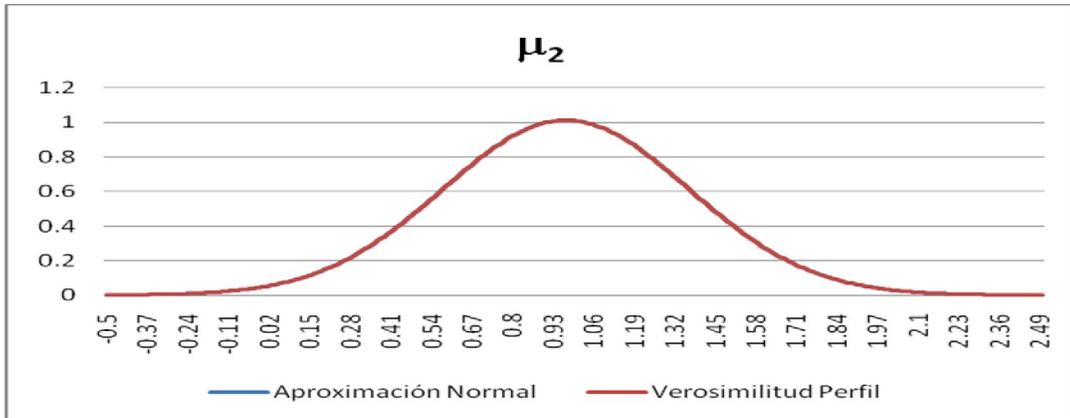
$$I_1 = \begin{pmatrix} .04505996 & .00027916 \\ .00027916 & .02015838 \end{pmatrix} \text{ y } I_2 = \begin{pmatrix} .15552895 & -.000032606 \\ -.000032606 & .02000326 \end{pmatrix}.$$

Como se puede ver las diagonales de las matrices observadas de Fisher por el método LVPLA toman valores muy parecidos a los elementos de la diagonal de la matriz

observada de Fisher obtenida por el método PLA, por lo que podemos ver que el método LVPLA es muy preciso, con respecto al método PLA.

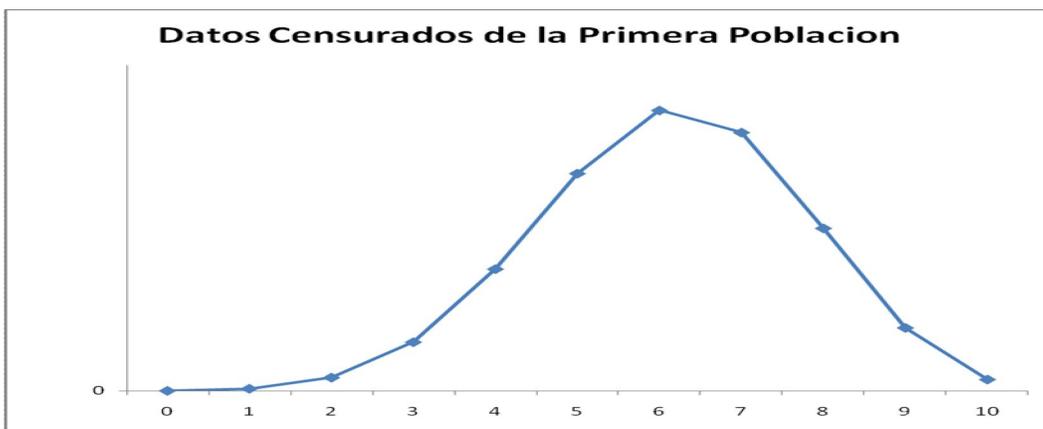
Si aproximamos cada una de las verosimilitudes perfiles calculadas numéricamente con la distribución normal, con media el estimador máximo verosímil y varianza la entrada correspondiente a la matriz observada de Fisher tenemos:





Como se ve en las anteriores gráficas la aproximación normal es muy buena para la verosimilitud perfil de cada uno de los parámetros de localización y escala.

Si graficamos la verosimilitud con respecto a los valores que puede tomar  $r_1$ , tenemos:



que como ya vimos se maximiza con  $r_1=6$ , que es el verdadero valor del número de datos censurados en la primera población.

Si deseamos hacer intervalos de confianza al 95% de cada uno de los parámetros tenemos lo siguiente:

$$\begin{aligned}\mu_1 &: [-.441633, .39047836], \varphi_1 : [-.224381, .33218137], \sigma_1^2 : [.63841803, 1.94325179] \\ \mu_2 &: [.20068852, 1.7466251], \varphi_2 : [.40067198, .95508883], \sigma_2^2 : [2.22853399, 6.75428866]\end{aligned}$$

Como se ve, todos los intervalos al 95% contienen el verdadero valor del parámetro, además los intervalos de confianza son muy parecidos a los obtenidos por el método PLA, lo cual nos confirma que el método LVPLA no pierde precisión con respecto al método PLA:

Para estimar la proporción de datos correspondientes a cada población utilizamos la siguiente fórmula:

$$\hat{\pi}_1 = \frac{r_1 + n_1}{N} = \frac{6 + 19}{50} = .5, \quad \hat{\pi}_2 = 1 - \hat{\pi}_1 = 1 - .5 = .5.$$

Podemos hacer inferencias de la proporción con una mezcla de Betas, de la siguiente manera:  $\text{Beta}(t+I, n-t)$  y  $\text{Beta}(t, n-t+I)$ , en este caso  $n=50$ ,  $t=25$ , entonces tenemos una mezcla de  $\text{Beta}(26, 25)$  y  $\text{Beta}(25, 26)$ . Si hacemos un intervalo de confianza para la proporción tenemos:  $[.367148, .638548]$

## **2.5 EJEMPLO CON TRES POBLACIONES**

Se generan ahora tres poblaciones normal con 15 datos cada uno. La primera población es  $N(0,1)$ , la segunda  $N(1,4)$  y la tercera  $N(0,4)$ .

N(0,1)	N(1,4)	N(0,4)
-1.7804	-0.1684	1.4124
-0.1097	0.1276	-2.2215
-0.2046	-1.5164	-0.2022
0.5126	3.2326	-3.4338
-0.2394	-0.4752	-0.5054
-1.6303	5.2145	-2.3126
-1.2753	-0.0123	1.1497
0.1312	2.4649	1.0186
-0.8207	2.8945	1.9711
-0.5279	-2.3543	0.6907
-0.1086	0.9353	1.0823
1.8653	-0.0965	-1.8324
1.2789	2.41262	4.0304
1.4782	0.6516	0.4357
-0.0525	1.12623	-1.2725

Decidimos tomar como región de censura el intervalo  $[2.5, \infty)$ , es decir, censura por la derecha, por lo cual nos quedamos con los siguientes datos:

N(0,1)	N(1,4)	N(0,4)
-1.7804	-0.1684	1.4124
-0.1097	0.1276	-2.2215
-0.2046	-1.5164	-0.2022
0.5126	-0.4752	-3.4338
-0.2394	-0.0123	-0.5054
-1.6303	2.4649	-2.3126
-1.2753	-2.3543	1.1497
0.1320	0.9353	1.0186
-0.8207	-0.0965	1.9711
-0.5279	2.41262	0.6907
-0.1086	0.6516	1.0823
1.8653	1.1262	-1.8324
1.27890		0.4357
1.4782		-1.2726
-0.0525		

Entonces tenemos:  $n_1=15$ ,  $n_2=12$  y  $n_3=14$ ,  $r=4$ , y la región de censura como  $[2.5, \infty)$ . Sea  $r_1$  y  $r_2$  el número de valores censurados de la población 1 y 2 respectivamente. Sabemos que  $r_3$ , el número de datos censurados de la población tres es  $r_3=4-r_1-r_2$ . Suponemos que conocemos los valores de  $r_1$ ,  $r_2$  y por lo tanto de  $r_3$ , sacamos los estimadores máximo verosímiles para los parámetros de localización y escala de cada una de las poblaciones.

Esto lo tenemos que hacer para cada una de los posibles valores que pueden tomar  $r_1$  y  $r_2$ , al mismo tiempo en este caso:

$$\binom{r+k-1}{r} = \binom{4+3-1}{4} = \binom{6}{4} = 15.$$

Valuamos cada una de las posibles  $r_1, r_2$  con sus respectivos parámetros de localización y escala, en la función de verosimilitud y nos quedamos con el que maximiza a ésta:

$r_1$	$r_2$	$r_3$	$\mu_1$	$\varphi_1$	$\mu_2$	$\varphi_2$	$\mu_3$	$\varphi_3$	L
0	0	4	-0.0988	0.0312	0.2380	0.2996	0.6506	0.8134	4.24281E-18
0	1	3	-0.0988	0.0312	0.4503	0.4253	0.4285	0.7530	6.12502E-18
0	2	2	-0.0988	0.0312	0.6534	0.5217	0.1953	0.6805	7.24588E-18
0	3	1	-0.0988	0.0312	0.9049	0.5918	-0.0502	0.5909	7.72844E-18
0	4	0	-0.0988	0.0312	1.0343	0.6638	-0.3092	0.4748	6.78533E-18
1	0	3	0.0935	0.2237	0.2380	0.2996	0.4285	0.7530	1.27641E-18
1	1	2	0.0935	0.2237	0.4503	0.4253	0.1953	0.6805	1.7939E-18
1	2	1	0.0935	0.2237	0.6534	0.5217	-0.0502	0.5909	2.03167E-18
1	3	0	0.0935	0.2237	0.9049	0.5918	-0.3092	0.4748	1.85733E-18
2	0	2	0.2795	0.3583	0.2380	0.2996	0.1953	0.6805	5.63188E-19
2	1	1	0.2795	0.3583	0.4503	0.4253	-0.0502	0.5909	7.57761E-19
2	2	0	0.2795	0.3583	0.6534	0.5217	-0.3092	0.4748	7.35568E-19
3	0	1	0.4592	0.4611	0.2380	0.2996	-0.0502	0.5909	2.97277E-19
3	1	0	0.4592	0.4611	0.4503	0.4253	-0.3092	0.4748	3.42828E-19
4	0	0	0.6327	0.5435	0.2380	0.2996	-0.30924	0.4748	1.5199E-19

Como se ve en la anterior tabla, los valores de  $r_1, r_2$  y  $r_3$  donde se maximiza la función de densidad es el vector (0,3,1), que son los verdaderos valores que se censuraron realmente.

Las matrices observadas de Fisher quedan como:

$$I_1 = \begin{pmatrix} .0709 & 0 \\ 0 & .0333 \end{pmatrix}, I_2 = \begin{pmatrix} .2310 & .0128 \\ .0128 & .04607 \end{pmatrix} \text{ y } I_3 = \begin{pmatrix} .2199 & .0030 \\ .0030 & .03701 \end{pmatrix}$$

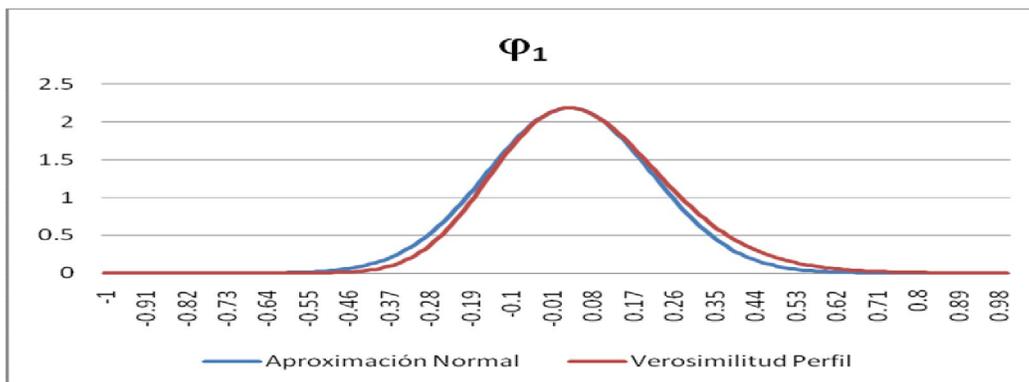
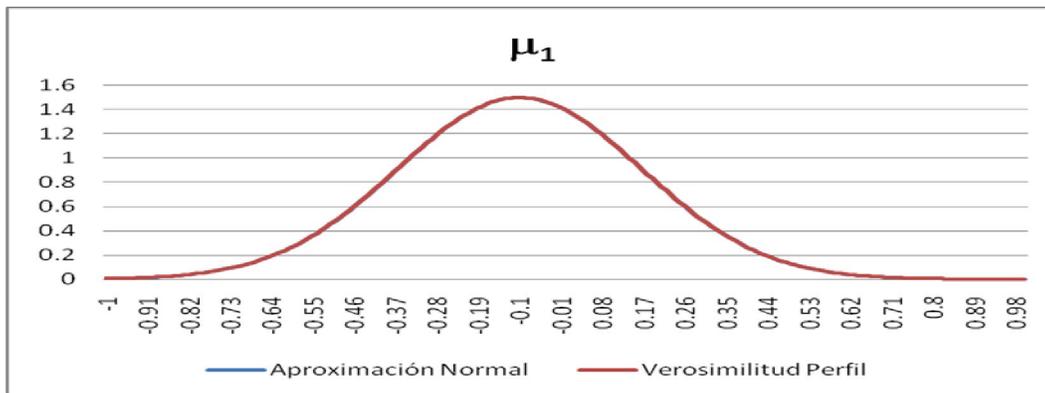
Los intervalos de confianza al 95% quedan como:

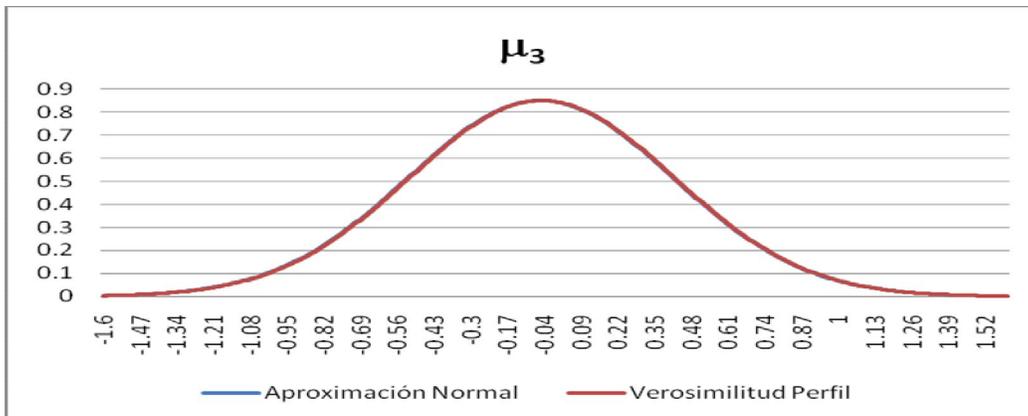
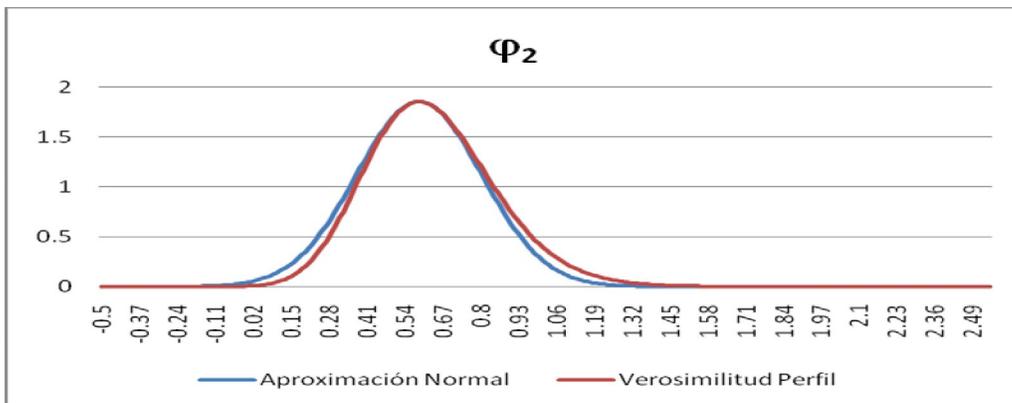
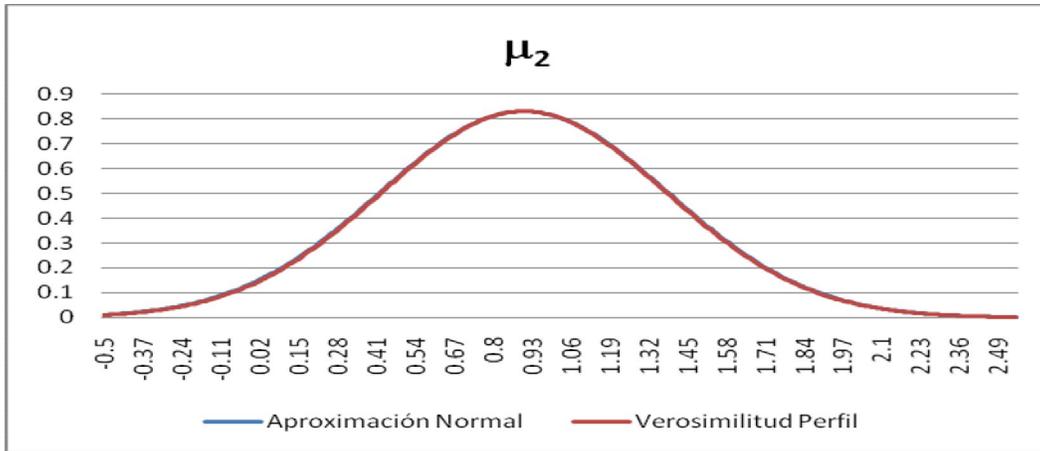
$\mu_1$ : [-.62093,.42328],  $\varphi_1$ : [-.32664,.38904], para  $\sigma_1^2$ : [.5203,2.1773]

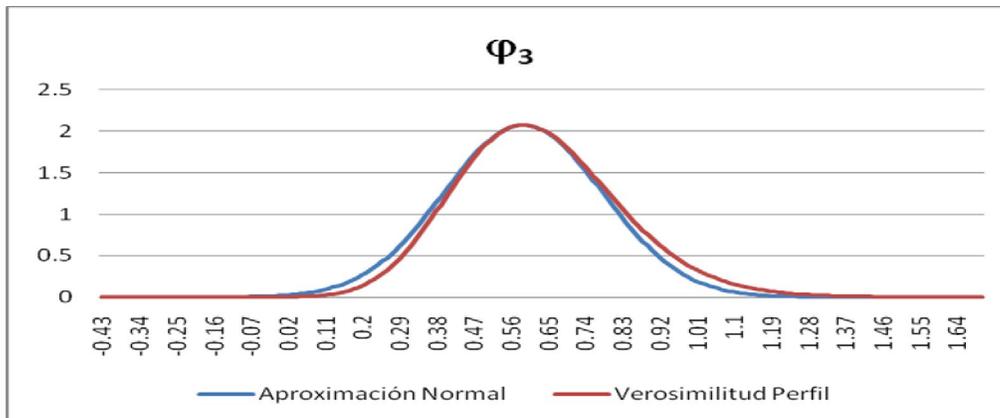
$\mu_2$ : [-.037,1.8468],  $\varphi_2$ : [.171141,1.01252], para  $\sigma_2^2$ : [1.4286,7.6934]

$\mu_3$ : [-1.0019,0.90148],  $\varphi_3$ : [.21375,.96795], para  $\sigma_3^2$ : [1.5334,6.9303].

Como ocurre en el caso anterior, todos los intervalos contienen el verdadero valor del parámetro. Si aproximamos cada verosimilitud perfil calculada de forma numérica con la distribución normal tenemos los siguientes resultados:







Como se ve en las anteriores gráficas, la aproximación no es excelente como lo fue en el primer ejemplo, esto es debido a que este ejemplo contenía menos datos, pero se tomó de este modo, para que hubiera menos datos censurados y se pudiera explicitar en la tabla todas las posibles combinaciones que se pueden dar de los datos censurados en las tres poblaciones. Aun así, la aproximación es buena.

Para estimar las proporciones tenemos:

$$\hat{\pi}_1 = \frac{15+0}{45} = .333, \quad \hat{\pi}_2 = \frac{12+3}{45} = .333 \text{ y } \hat{\pi}_3 = 1 - \hat{\pi}_1 - \hat{\pi}_2 = .333.$$

Para hacer inferencias sobre las proporciones utilizamos nuevamente la mezcla de Betas, para  $\pi_1$  utilizamos Beta(15.5,30.5), para  $\pi_2$  Beta(12.5,44.5).

Los intervalos de confianza quedan:  $\pi_1$ : [.2093,.4780],  $\pi_2$ : [.14061,.3907].

Si lo hubiéramos hecho por el método PLA, los estimadores máximos verosímiles quedan:

$$\hat{\mu}_1 = -0.0763, \quad \hat{\varphi}_1 = 0.058, \quad \hat{\mu}_2 = 0.718, \quad \hat{\varphi}_2 = 0.5232, \quad \hat{\mu}_3 = 0.1378, \quad \hat{\varphi}_3 = 0.6642, \\ \hat{\pi}_1 = .3359, \quad \hat{\pi}_2 = .3129.$$

En este caso la matriz de Fisher es de tamaño ocho. Por efectos de espacio, no se va a poner toda la matriz de Fisher, únicamente la diagonal que es donde están las varianzas:

$$\hat{V}(\hat{\mu}_1) = .0800, \quad \hat{V}(\hat{\varphi}_1) = .0411, \quad \hat{V}(\hat{\mu}_2) = .43103, \quad \hat{V}(\hat{\varphi}_2) = .07925, \quad \hat{V}(\hat{\mu}_3) = 0.50377, \\ \hat{V}(\hat{\varphi}_3) = 0.0732, \quad \hat{V}(\hat{\pi}_1) = .00503, \quad \hat{V}(\hat{\pi}_2) = .0072.$$

En este caso las varianzas de los estimadores por el método PLA son un poco más grandes que las varianzas de los estimadores con el método LVPLA, esto es debido al que el método PLA, no es tan preciso cuando tenemos una matriz de dimensión más grande, ya que surgen problemas numéricos y computacionales.

## **2.6 SIMULACIONES DE LOS DOS EJEMPLOS ANTERIORES**

Se hicieron simulaciones para los dos ejemplos anteriores. Para el primer ejemplo se generaron 300 muestras de tamaño 25 provenientes de una Normal (0,1) y 300 muestras de tamaño 25 de una Normal (1,4). Para cada uno de estas muestras se censuran las observaciones que cayeron en el intervalo [.2,1], y se estimaron los parámetros de localización y escala como se hizo en el ejemplo anterior, y se obtuvo lo siguiente:

Parámetro	Porcentaje de Intervalos que no contienen al verdadero valor
$\mu_1$	4.66%
$\varphi_1$	6%
$\mu_2$	3.33%
$\varphi_2$	4%

Se realizó lo mismo para el tercer ejemplo, generando 300 muestra de tamaño 15 provenientes de una Normal (0,1), N(1,4) y N(0,4), se censuraron los datos mayores a 2.5, y se estimaron los parámetros a través del método LVPLA, como se hizo en el ejemplo 3 y se obtuvo:

Parámetro	Porcentaje de Intervalos que no contienen al verdadero valor
$\mu_1$	7.33%
$\varphi_1$	2.66%
$\mu_2$	4.33%
$\varphi_2$	6%
$\mu_3$	3.33%
$\varphi_3$	4.66%

Como se ve en las anteriores tablas, se aproximan al 5% de los intervalos que no contienen el verdadero valor del parámetro.

Por último, aunque bajo el método LVPLA, se tienen que considerar  $\binom{r+k-1}{r}$  casos, no se tiene que maximizar cada uno de estos, ya que al trabajar cada población de forma independiente, en cada una de ellas se deben contemplar  $r+1$  casos, con  $i=0, \dots, r$ , y como el número de poblaciones es  $k$ , debemos encontrar los estimadores máximo verosímiles  $k(r+1)$ . Los cuales con una  $r$  y  $k$  grandes son menores que el anterior número. Aunque para construir la verosimilitud sí se tienen que considerar los  $\binom{r+k-1}{r}$  casos.

El método LVPLA es más eficiente computacionalmente que el método PLA, ya que el segundo método trabaja con matrices de tamaño  $3k-1$ , mientras que el método LVPLA trabaja con  $k$  matrices de tamaño 2, que son más fáciles de calcular numéricamente las inversas que la matriz de tamaño  $3k-1$ , y como se vió en los ejemplos anteriores el método LVPLA, tiene una precisión muy buena con respecto al método PLA.

### **3. MANEJO DE DATOS PERTENECIENTES A “K”**

#### **POBLACIONES BAJO CENSURA ALEATORIA PARCIALMENTE**

#### **IDENTIFICADOS**

El propósito de este capítulo consiste en realizar una extensión del método Latent Variable Profile Likelihood Analysis (LVPLA), para el caso en que los datos de la mezcla de las poblaciones se encuentren bajo censura aleatoria. Esta extensión es un resultado desarrollado en el presente documento y es la principal aportación de éste.

Supongamos que tenemos datos de “ $k$ ” poblaciones, todas ellas pertenecientes a una familia de localización y escala. Estos datos, al igual que en el contexto de la sección anterior, también están censurados. La diferencia, ahora, consiste en que los datos están censurados de forma aleatoria, es decir que no se define previamente una región de censura; de hecho la región de censura no existe, ya que cada dato puede ser censurado o no, dependiendo si el factor o factores que provocan la censura ocurrieron antes (si es censura por la derecha) o después (si es censura por la izquierda) de haber podido realizar la observación del dato. Esto es diferente del caso anterior donde se establece previamente la región de censura y se cuentan como censurados los datos que caen en dicha región.

Nuevamente suponemos que los datos de las “ $k$ ” poblaciones provienen cada uno de una muestra aleatoria.

Entonces, tenemos la siguiente información:

- 1) Si el dato no fue censurado, además del valor de la observación, se registra la población a la que pertenece. Esto es igual que en el caso de censura no aleatoria.
- 2) Si el dato fue censurado aleatoriamente, entonces no contamos con el verdadero valor de la observación, ni la población a la que pertenece, pero si sabemos de la existencia de dicha observación, además de que bajo la censura aleatoria se registra el tiempo en que fue censurado, por lo que sabemos el mínimo o máximo valor que pudo haber tomado la observación, dependiendo si es

censura aleatoria por la derecha o por la izquierda. Por ejemplo, si la censura es por la derecha y tenemos un dato censurado, lo que sabemos es el tiempo en que la censura ocurrió, por lo que la observación “ocurriría” en un momento posterior al que tenemos registrado como censura.

El propósito de este capítulo y, el central de este trabajo es estimar los parámetros de localización y escala para cada uno de las poblaciones, a través de la asignación de los datos censurados a una población específica.

Lo anterior se va hacer por medio de los dos métodos anteriores: PLA, y más específicamente el método LVPLA, aplicados en el caso de censura aleatoria.

Un ejemplo donde esta situación se aplicaría es cuando tenemos un grupo de pacientes que padecen de diabetes, los cuales pueden morir por muchas causas. Cada uno de los pacientes puede abandonar el estudio por diversas causas de forma independiente a los demás pacientes, por lo que no registramos su tiempo de muerte, ni la causa de ésta, sino únicamente su tiempo de abandono del estudio, por otra parte si el paciente no abandona el estudio y muere, además de registrar el tiempo de muerte registramos la causa, aquí las poblaciones corresponden a las distintas causas de muerte.

Definimos nuevamente a  $f_i(x, \mu_i, \sigma_i)$  y  $F_i(x, \mu_i, \sigma_i)$  como las funciones de densidad y distribución de la  $i$ 'ésima población, cada una con parámetros de localización  $\mu_i$ , y escala  $\sigma_i$ , y sea  $\pi_i$ , la proporción de datos que pertenecen a la  $i$ 'ésima población. Ya sabemos que dichos valores deben cumplir con lo siguiente:

$$1) \pi_i \geq 0, \quad i=1, \dots, k.$$

$$2) \sum_{j=1}^k \pi_j = 1.$$

La distribución “mezcla” queda definida, al igual que en el caso anterior, como:

$$\sum_{i=1}^k \pi_i F_i(x, \mu_i, \sigma_i).$$

Para la censura aleatoria además de las funciones de densidad y distribución del tiempo de falla para cada una de las “ $k$ ” poblaciones, necesitamos definir las funciones de densidad y distribución del tiempo de censura, los cuales los denotaremos como  $g(t)$  y  $G(t)$ , respectivamente. Es importante mencionar que supondremos que la distribución del tiempo de censura es independiente de la distribución del tiempo de fallas, es la misma para cada población y que la distribución de censura no contiene parámetros de interés.

Supongamos que tenemos  $n_i$  observaciones registradas de la  $i$ 'ésima población, la suma de las  $n_i$ 's la llamamos  $n$ . Denotamos a  $r$  como el número de observaciones censuradas de las cuales no sabemos su tiempo de falla, ni a la población que pertenecen. El número total de datos es  $N$ , por lo que  $N=n+r$ .

Utilizamos una variable adicional llamada  $\delta$ , que toma el valor 1 si el dato fue observado y clasificado en una población y 0 en caso de que hubiera sido censurado.

Sea  $x_{ij}$  la  $j$ 'ésima observación registrada de la  $i$ 'ésima población, en este caso  $\delta_{ij}$  es igual a 1. Por otra parte, si  $t_s$ , es la  $s$ 'ésima observación censurada se tiene que  $\delta_s$  toma el valor 0.

Vamos a desarrollar la metodología para el caso en que la censura sea por la derecha, ya que para los demás casos quedaría de forma similar.

Como la censura es por la derecha, para cada dato se registrará lo que ocurra primero: el tiempo de muerte o el tiempo de censura, es decir:

- 1) Si la observación que proviene de la mezcla de las distribuciones se dio antes de la causa de censura, se identifica el valor de la observación y además se identifica la población a la que pertenece.

- 2) Si la causa de censura se dio antes que la observación proveniente de la mezcla, registraremos el tiempo de censura, no el de la observación que ocurriría después, ni tampoco la población a la que pertenecería.

Como en el contexto anterior es necesario construir la función de verosimilitud, empezaremos con los datos que fueron registrados:  $x_{ij}$  con  $i=1, \dots, k, j=1, \dots, n_i$ , tenemos que parte de la verosimilitud queda como:

$$(1 - G(x_{ij}))\pi_i f_i(x_{ij}, \mu_i, \sigma_i).$$

Para los datos censurados, sabemos el tiempo en que ocurrió la censura, y que la observación de la mezcla es mayor a  $t_s$ , por lo que la verosimilitud tiene que considerar lo anterior, es decir esa parte queda como:

$$g(t_s)(1 - \sum_{i=1}^k \pi_i F_i(t_s, \mu_i, \sigma_i))$$

Si juntamos ambas partes tomando en cuenta la independencia tenemos:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{X}) \propto \prod_{s=1}^r g(t_s)(1 - \sum_{i=1}^k \pi_i F_i(t_s, \mu_i, \sigma_i)) \left\{ \prod_{i=1}^k \prod_{j=1}^{n_i} \pi_i (1 - G(x_{ij})) f_i(x_{ij}, \mu_i, \sigma_i) \right\}$$

Podemos agrupar la verosimilitud con los datos que involucran la distribución de censura y las distribuciones de fallas:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{X}) \propto \left\{ \prod_{i=1}^k \prod_{j=1}^{n_i} \pi_i f_i(x_{ij}, \mu_i, \sigma_i) \right\} \left\{ \prod_{s=1}^r (1 - \sum_{i=1}^k \pi_i F_i(t_s, \mu_i, \sigma_i)) \right\} \prod_{s=1}^r g(t_s) \left\{ \prod_{i=1}^k \prod_{j=1}^{n_i} (1 - G(x_{ij})) \right\}$$

Como mencionamos anteriormente, la función de densidad de la censura no contiene parámetros de interés, por lo cual podemos prescindir de estos:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{X}) \propto \prod_{i=1}^k \prod_{j=1}^{n_i} \pi_i f_i(x_{ij}, \mu_i, \sigma_i) \left\{ \prod_{s=1}^r (1 - \sum_{i=1}^k \pi_i F_i(t_s, \mu_i, \sigma_i)) \right\}$$

Una vez obtenida la función de verosimilitud podemos utilizar los métodos PLA y LVPLA para poder encontrar los estimadores máximo verosímiles de los parámetros.

### **3.1 PROFILE LIKELIHOOD ANALYSIS (PLA)**

Como en el caso anterior, el método Profile Likelihood Analysis (PLA) encuentra los estimadores máximo verosímiles a través de la verosimilitud.

Primero obtenemos la función de log-verosimilitud:

$$l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x}) = \sum_{s=1}^r \ln\left(1 - \sum_{i=1}^k \pi_i F_i(t_s, \mu_i, \sigma_i)\right) + \sum_{i=1}^k \sum_{j=1}^{n_i} \ln(f_i(x_{ij}, \mu_i, \sigma_i)) + \sum_{i=1}^k n_i \ln(\pi_i).$$

Después encontramos las derivadas parciales con respecto a los parámetros de interés:

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x})}{\partial \mu_i} = -\sum_{s=1}^r \frac{1}{\left(1 - \sum_{j=1}^k \pi_j F_j(t_s, \mu_j, \sigma_j)\right)} \frac{\partial F_i(t_s, \mu_i, \sigma_i)}{\partial \mu_i} + \sum_{j=1}^{n_i} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \mu_i}$$

$i=1, 2, \dots, k.$

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x})}{\partial \sigma_i} = -\sum_{s=1}^r \frac{1}{\left(1 - \sum_{j=1}^k \pi_j F_j(t_s, \mu_j, \sigma_j)\right)} \frac{\partial F_i(t_s, \mu_i, \sigma_i)}{\partial \sigma_i} + \sum_{j=1}^{n_i} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \sigma_i}$$

$i=1, 2, \dots, k.$

Para los parámetros de proporción tenemos:

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{X})}{\partial \pi_i} = \frac{-F_i(t_s, \mu_i, \sigma_i)}{(1 - \sum_{j=1}^k \pi_j F_j(t_s, \mu_j, \sigma_j))} + \frac{n_i}{\pi_i}, \quad i=1, 2, \dots, k-1.$$

Tenemos  $k$  parámetros de localización,  $k$  parámetros de escala y  $k-1$  proporciones que estimar. Cada una de las derivadas parciales se iguala a cero para obtener los estimadores máximo verosímiles. Como tenemos un sistema de  $3k-1$  ecuaciones con  $3k-1$  incógnitas, al igual que en el caso anterior este sistema se puede complicar mucho, además que por ser censura aleatoria, las expresiones son más difíciles que en el caso de censura no aleatoria.

Si quisiéramos hacer inferencias sobre los parámetros tenemos que calcular la matriz observada de Fisher, lo cual involucraría obtener la matriz de segundas derivadas parciales, la cual es de tamaño  $3k-1$ , y sacarle inversa a dicha matriz. Nuevamente esto es muy complicado ya que el tamaño de la matriz puede crecer mucho con una  $k$  no muy grande.

En el caso de que la censura fuese por la izquierda, la función de verosimilitud quedaría de la siguiente forma:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{X}) \propto \left\{ \prod_{s=1}^r \left( \sum_{i=1}^k \pi_i F_i(t_s, \mu_i, \sigma_i) \right) \right\} \left\{ \prod_{i=1}^k \prod_{j=1}^{n_i} \pi_i f_i(x_{ij}, \mu_i, \sigma_i) \right\} \prod_{s=1}^r g(t_s, \mu, \sigma) \left\{ \prod_{i=1}^k \prod_{j=1}^{n_i} G(x_{ij}, \mu, \sigma) \right\}$$

y por lo tanto la log-verosimilitud es:

$$l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{X}) = \sum_{s=1}^r \ln \left( \sum_{i=1}^k \pi_i F_i(t_s, \mu_i, \sigma_i) \right) + \sum_{i=1}^k \sum_{j=1}^{n_i} \ln(f_i(x_{ij}, \mu_i, \sigma_i)) + \sum_{i=1}^k n_i \ln(\pi_i)$$

Por el método PLA, las derivadas parciales quedan como:

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{X})}{\partial \mu_i} = \sum_{s=1}^r \frac{1}{\left( \sum_{j=1}^k \pi_j F_j(t_s, \mu_j, \sigma_j) \right)} \frac{\partial \pi_i F_i(t_s, \mu_i, \sigma_i)}{\partial \mu_i} + \sum_{j=1}^{n_i} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \mu_i}$$

$i=1, 2, \dots, k.$

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x})}{\partial \sigma_i} = \sum_{s=1}^r \frac{1}{\left(\sum_{j=1}^k \pi_j F_j(t_s, \boldsymbol{\mu}_j, \sigma_j)\right)} \frac{\partial \pi_i F_i(t_s, \boldsymbol{\mu}_i, \sigma_i)}{\partial \sigma_i} + \sum_{j=1}^{n_i} \frac{1}{f_i(x_{ij}, \boldsymbol{\mu}_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \boldsymbol{\mu}_i, \sigma_i)}{\partial \sigma_i}$$

$i=1, 2, \dots, k.$

Para los parámetros de proporción tenemos:

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}, \mathbf{x})}{\partial \pi_i} = \sum_{s=1}^r \frac{F_i(t_s, \boldsymbol{\mu}_i, \sigma_i)}{\left(\sum_{j=1}^k \pi_j F_j(t_s, \boldsymbol{\mu}_j, \sigma_j)\right)} + \frac{n_i}{\pi_i} \quad i=1, 2, \dots, k-1$$

A continuación vamos a ver una extensión del método LVPLA en el caso de censura aleatoria.

### **3.2 LATENT VARIABLE PROFILE LIKELIHOOD ANALYSIS (LVPLA)**

En el caso de censura no aleatoria el método Latent Variable Profile Likelihood Analysis (LVPLA), suponemos que conocemos el valor de  $r_i$ , es decir, el número de datos censurados que pertenecen a la  $i$ 'ésima población.

Esto se podía hacer ya que la región de censura es la misma para todas las observaciones. En el caso de censura aleatoria no existe una región de censura para todas las poblaciones, ya que cada dato puede ser censurado de forma independiente, por lo que podemos decir que cada dato tiene su propia región de censura, que no necesariamente es ajena a la región de censura de las demás observaciones, ni con respecto a la región de no censura, por lo que no podemos agrupar las distintas observaciones censuradas.

Lo que hace el método LVPLA para datos censurados de forma aleatoria es suponer que para cada uno de los datos censurados conocemos la población a la que pertenece. Esto

lo hacemos a través de otras variables latentes denotadas por  $\lambda_{si}$  vale 1 si la  $s$ 'ésima observación censurada pertenece a la  $i$ 'ésima población, y 0 en otro caso. Como las observaciones censuradas pertenecen a una y solamente una de las  $k$  poblaciones entonces:

$$\sum_{i=1}^k \lambda_{si} = 1,$$

$$\sum_{s=1}^r \sum_{i=1}^k \lambda_{si} = r$$

Al hacer esto, la verosimilitud queda de la siguiente forma:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{x}) \propto \prod_{i=1}^k \prod_{s=1}^r (1 - F_i(t_s, \mu_i, \sigma_i))^{\lambda_{si}} \prod_{i=1}^k \prod_{j=1}^{n_i} f_i(x_{ij}, \mu_i, \sigma_i).$$

La cual se puede expresar de la siguiente forma:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{x}) \propto \prod_{i=1}^k \left\{ \prod_{s=1}^r (1 - F_i(t_s, \mu_i, \sigma_i))^{\lambda_{si}} \prod_{j=1}^{n_i} f_i(x_{ij}, \mu_i, \sigma_i) \right\}.$$

que es equivalente a:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{x}) \propto \prod_{i=1}^k L_i(x_i, \mu_i, \sigma_i).$$

Donde  $L_i(\mu_i, \sigma_i, x_i)$  son las verosimilitudes perfil para cada una de las " $k$ " poblaciones.

$$L_i(\mu_i, \sigma_i, x_i) \propto \prod_{s=1}^r (1 - F_i(t_s, \mu_i, \sigma_i))^{\lambda_{si}} \prod_{j=1}^{n_i} f_i(x_{ij}, \mu_i, \sigma_i). \quad i=1, 2, \dots, k.$$

Al hacer esto podemos manejar de manera separada cada una de las verosimilitudes perfil y encontrar los estimadores máximo verosímiles, para ello, obtenemos la función de log-verosimilitud:

$$l_i(\mu_i, \sigma_i, x_s) = \sum_{s=1}^r \lambda_{si} \ln(1 - F_i(t_s, \mu_i, \sigma_i)) + \sum_{j=1}^{n_j} \ln(f_i(x_{ij}, \mu_i, \sigma_i)).$$

Las derivadas parciales quedan de la siguiente forma:

$$\frac{\partial l_i(\mu_i, \sigma_i, x)}{\partial \mu_i} = \sum_{s=1}^r \frac{-\lambda_{si}}{(1 - F_i(t_s, \mu_i, \sigma_i))} \frac{\partial F_i(t_s, \mu_i, \sigma_i)}{\partial \mu_i} + \sum_{j=1}^{n_j} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \mu_i}.$$

$i=1, 2, \dots, k.$

$$\frac{\partial l_i(\mu_i, \sigma_i, x)}{\partial \sigma_i} = \sum_{s=1}^r \frac{-\lambda_{si}}{(1 - F_i(t_s, \mu_i, \sigma_i))} \frac{\partial F_i(t_s, \mu_i, \sigma_i)}{\partial \sigma_i} + \sum_{j=1}^{n_j} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \sigma_i}.$$

$i=1, 2, \dots, k.$

En este caso tenemos “ $k$ ” sistemas de dos ecuaciones con dos incógnitas, los cuales son más fáciles de resolver que un sistema de  $3k-1$  ecuaciones con  $3k-1$  incógnitas.

Una vez resueltos estos sistemas tenemos los estimadores máximo verosímiles para los parámetros de localización y escala, el inconveniente es que estamos suponiendo que conocemos a que población pertenece cada una de los datos censurados, lo cual no es así, por lo que lo anterior lo hacemos para todas las combinaciones posibles de los datos censurados sobre las poblaciones. Cada una de los “ $r$ ” datos censurados puede pertenecer a las  $k$  poblaciones, por lo que, el número de combinaciones posibles es  $k^r$ .

Valuamos la función de verosimilitud en el vector de parámetros de localización y escala maximizados, para todas y cada una de las combinaciones posibles. Nos

quedamos con la combinación en la cual se maximiza la verosimilitud. Los estimadores que utilizan son los asociados a dicha combinación.

Obtenemos las segundas derivadas parciales evaluadas en los estimadores previamente escogidos y con estos construimos  $k$  matrices de Fisher. Estas  $k$  matrices de Fisher de tamaño 2 son más fáciles de calcular que si manejara una sola matriz de Fisher de tamaño  $3k-1$ , que es lo que se hace con el método PLA.

En el caso de que la censura sea por la izquierda tenemos para el método LVPLA, que la función de verosimilitud se puede escribir como:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{x}) \propto \prod_{i=1}^k \prod_{s=1}^r (F_i(t_s, \mu_i, \sigma_i))^{\lambda_{si}} \prod_{i=1}^k \prod_{j=1}^{n_i} f_i(x_{ij}, \mu_i, \sigma_i).$$

En caso de que utilicemos el método LVPLA, tenemos que las expresiones para las derivadas parciales con respecto a los parámetros de localización y escala, para cada una de las  $k$  log-verosimilitudes quedan como:

$$\frac{\partial l_i(\mu_i, \sigma_i, \mathbf{x})}{\partial \mu_i} = \sum_{s=1}^r \frac{\lambda_{si}}{F_i(\mu_i, \sigma_i)} \frac{\partial F_i(t_s, \mu_i, \sigma_i)}{\partial \mu_i} + \sum_{j=1}^{n_j} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \mu_i}, i=1, 2, \dots, k$$

$$\frac{\partial l_i(\mu_i, \sigma_i, \mathbf{x})}{\partial \sigma_i} = \sum_{s=1}^r \frac{\lambda_{si}}{F_i(\mu_i, \sigma_i)} \frac{\partial F_i(t_s, \mu_i, \sigma_i)}{\partial \sigma_i} + \sum_{j=1}^{n_j} \frac{1}{f_i(x_{ij}, \mu_i, \sigma_i)} \frac{\partial f_i(x_{ij}, \mu_i, \sigma_i)}{\partial \sigma_i}, i=1, 2, \dots, k.$$

El método PLA calcula de forma directa la proporción de datos pertenecientes a cada población. El método LVPLA, en caso de que los datos hayan sido censurados de forma no aleatoria, estima el número de datos censurados que pertenece a cada población y con esto se puede obtener una estimación para la proporción de cada población.

Para el caso de datos censurados aleatoriamente, primero obtenemos el número de datos censurados para cada población de la siguiente forma:

$$r_i = \sum_{s=1}^r \lambda_{si}, \text{ con } i=1, \dots, k.$$

Una vez encontrada la estimación del número de datos censurados pertenecientes a cada población podemos estimar la proporción como lo hacíamos anteriormente:

$$\hat{\pi}_i = \frac{r_i + n_i}{N}, \text{ con } i=1, 2, \dots, k$$

Para hacer inferencias sobre las proporciones utilizamos una mezcla de Betas, la cual se describió en la sección anterior.

### **3.3 EJEMPLO CON DOS POBLACIONES**

Se generan 25 datos provenientes de la Normal(0,1), otros 25 datos de la Normal(1,4). Como la censura es aleatoria, no podemos escoger una región de censura, sino que se debe censurar aleatoriamente, esto lo hacemos generando 50 datos Normales(1.7,4), las cuales van a censurar a las dos poblaciones, quedándonos con el mínimo de ellas. La información inicial que tenemos es:

Población 1	Población 2	Censura	
1.5882	-0.5301	3.4039	3.1349
-1.2250	2.6513	2.1981	0.5927
-1.1767	2.1015	1.2378	2.1964
-0.7401	4.1699	1.7810	5.6595
-1.9754	-0.7918	1.3630	2.2673
-1.0284	1.0689	3.0325	3.9210
-1.3921	-0.6653	3.4416	5.0066
-0.4959	-1.4926	1.8341	2.0482
-0.8564	1.3919	0.5575	2.5141
-0.1140	0.6422	0.9121	0.3316
0.1253	0.8632	4.9383	0.3077
-0.5587	3.2246	3.9283	3.7823
0.2181	-1.0703	2.4968	1.7402

1.2602	-0.1977	-1.2697	3.2765
0.3127	0.2881	2.5478	3.2291
-0.9648	-0.8469	-0.1247	0.2991
0.7795	0.7158	-0.6027	0.0007
-0.1820	4.9639	1.2863	1.3093
0.8894	0.7607	1.9030	0.3564
0.8891	0.4728	-1.5237	5.0770
-0.0542	-0.9634	1.6649	-0.3619
0.2277	0.6188	3.8687	2.5068
-1.1618	0.4003	-1.3837	0.4876
1.0328	0.9825	1.4865	2.5815
-0.2153	-2.2225	2.0285	0.4284

Si, ahora, nos quedamos con el mínimo entre la población y la censura tenemos lo siguiente:

Población 1	Población 2	Dato Censurado
1.5882	-0.5301	-1.2697
-1.2250	2.1015	-0.6027
-1.1767	4.1699	-1.5237
-0.7401	-0.7918	-1.3837
-1.9754	1.0689	0.5927
-1.0284	-0.6653	0.3316
-1.3921	-1.4926	1.3077
-0.4959	1.3919	0.0007
-0.8564	3.2246	0.3093
-0.1140	-1.0703	0.3564
0.1253	-0.1977	
-0.5587	0.2881	
0.2181	-0.8469	
0.3127	0.4728	
-0.9648	-0.9634	
-0.1820	0.6188	
0.8894	0.4003	
-0.0542	0.9825	
0.2277	-2.2225	
1.0328		
-0.2153		

Entonces tenemos 21 datos que sabemos que pertenecen a la población 1, 19 datos que pertenecen a la población 2, y 10 datos que están censurados, de los cuales no sabemos su verdadero valor ni la población a la que pertenecen, pero si sabemos el tiempo de censura, lo cual significa que el valor de la observación es mayor al dato que registramos como censurado.

Si lo decidimos resolver por el método Profile Likelihood Analysis tenemos que la log-verosimilitud queda como:

$$l(x, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\pi}) = \sum_{s=1}^r \ln(1 - \sum_{i=1}^k \pi_i \Phi_i((t_s - \mu_i)e^{-\varphi_i})) + \sum_{i=1}^k \sum_{j=1}^{n_i} \left( \frac{1}{2} (x_{ij} - \mu_i)^2 e^{-2\varphi_i} - n\varphi_i \right) + \sum_{i=1}^k n_i \ln(\pi_i)$$

A continuación se obtienen las derivadas parciales, se igualan a cero y se tiene un sistema de cinco ecuaciones con cinco incógnitas. Al hacer lo anterior, los estimadores máximos verosímiles quedan como:

$$\hat{\pi}_1 = 0.4926, \quad \hat{\mu}_1 = -0.2034, \quad \hat{\varphi}_1 = -0.1279, \quad \hat{\mu}_2 = 0.6281, \quad \hat{\varphi}_2 = 0.4545$$

La matriz observada de Fisher valuada en los anteriores parámetros queda como:

$$I = \begin{pmatrix} .006204 & .00207 & .00083 & -.00216 & .00006 \\ .00207 & .038886 & .004783 & -.004822 & -.000152 \\ .00083 & .004783 & .025448 & -.00245407 & -.0002313 \\ -.00216 & -.004822 & -.00245407 & .1187028 & .00858041 \\ .00006 & -.000152 & -.0002313 & .00858041 & .0259484 \end{pmatrix}$$

Los intervalos de confianza al 95% son:

$$\pi_1 : (.33826, .64703), \quad \mu_1 : (-.5898, .18299), \quad \varphi_1 : (-.4406, .18472), \quad \mu_2 : (-.472, 1.303), \\ \varphi_2 : (.13876, .7702).$$

Aplicando la parametrización inversa para encontrar los intervalos de las varianzas poblacionales tenemos:

$$\sigma_1^2 : (.4142, 1.4469) \text{ y } \sigma_2^2 : (1.3198, 4.6665).$$

Para resolver el problema con el método PLA, tuvieron que hacerse cálculos muy complejos, además al obtener la matriz de Fisher de tamaño 5 puede ser complicado numéricamente y esto puede generar imprecisión al momento de obtener la inversa del Hessiano, a continuación se va a hacer por el método LVPLA.

Al hacerlo por el método LVPLA, suponemos que sabemos a qué población pertenecen cada uno de los datos censurados, como tenemos 2 poblaciones y 10 datos censurados, el número total de combinaciones posibles es de:  $2^{10}=1024$ .

Calculamos los estimadores máximos verosímiles en los 1024 casos, valuamos las verosimilitudes perfil y con esto podemos valorar la verosimilitud en cada uno de los casos, quedándonos con aquel que maximizó la verosimilitud, en este caso tenemos que los estimadores máximos verosímiles con este método son:

$$\hat{\mu}_1 = -0.26999, \hat{\varphi}_1 = -0.182856, \hat{\mu}_2 = 0.680933, \hat{\varphi}_2 = 0.467535$$

Mediante este método tenemos dos matrices observadas de Fisher:

$$I_1 = \begin{pmatrix} 0.0312177 & 0.00201649 \\ 0.00201649 & 0.0226181 \end{pmatrix} \text{ y } I_2 = \begin{pmatrix} 0.11383 & .00626433 \\ .00626433 & .026521 \end{pmatrix}$$

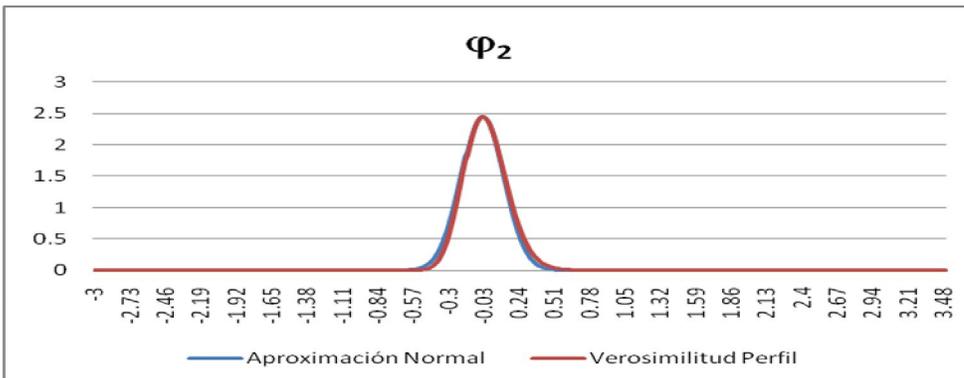
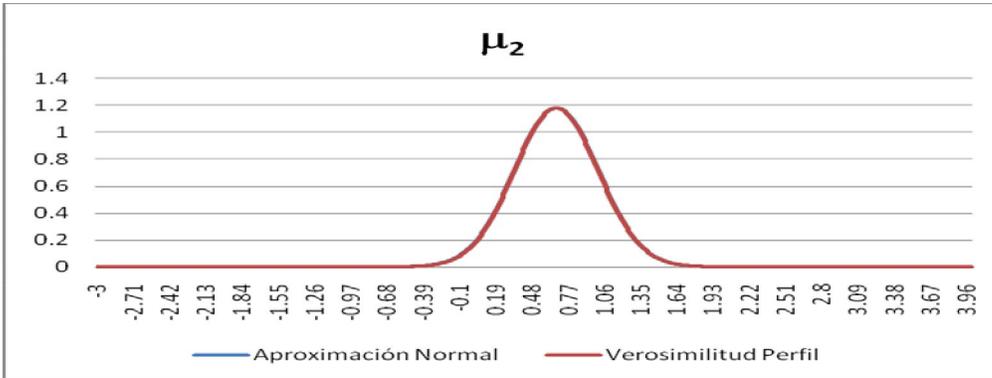
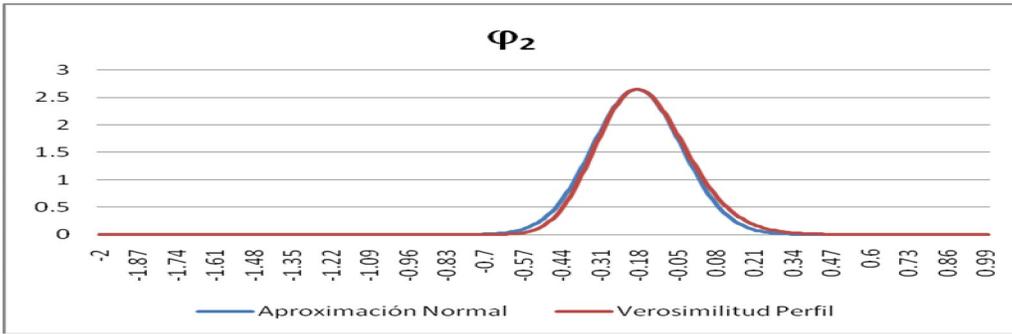
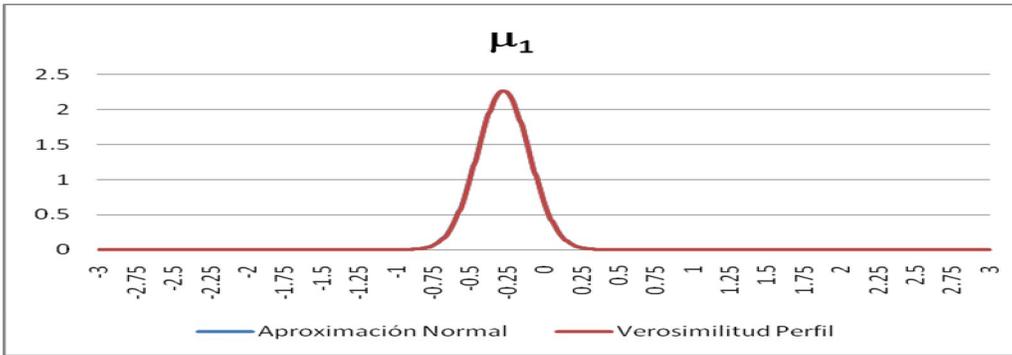
Los intervalos de confianza son:

$$\mu_1 : (-.6162, .0763), \varphi_1 : (-.4776, .1119), \mu_2 : (.019656, 1.342213), \varphi_2 : (.48344, .786727).$$

Los intervalos de confianza para la varianza quedan como:

$$\sigma_1^2 : (.3847, 1.25) \text{ y } \sigma_2^2 : (1.3453, 4.8232).$$

Se graficamos las verosimilitudes perfil, con las respectivas aproximaciones normal, tenemos:



Como se puede apreciar en las cuatro gráficas la aproximación normal es muy buena para las verosimilitudes perfil.

Si quisiéramos estimar la proporción de datos que pertenecen a cada población, primero tendríamos que estimar el número de datos censurados correspondiente a cada población. En el caso de censura aleatoria es necesario ver donde es asignado dato por dato.

Dato	Población Asignada	Población Verdadera
-1.2698	1	1
-0.6027	1	1
-1.5237	1	1
-1.3837	1	1
0.5927	2	2
0.3317	2	2
1.3077	2	2
0.0007	2	2
0.3093	2	2
0.3564	2	2

En este caso, todos los datos censurados fueron asignados de forma correcta, esto no siempre se va a dar. Como se puede apreciar en la anterior gráfica, lo que hace este método es separar en regiones los datos censurados, es decir, que para los datos de menor valor se les asigna la población 1, mientras que los datos de mayor valor la población 2. En caso de que hubiéramos tenido un dato de menor valor en la población 2, éste hubiera sido asignado a la población 1.

Los estimadores para el número de datos censurados de la población 1 y 2 son respectivamente:

$$r_1 = 4 \text{ y } r_2 = 6.$$

Las proporciones para cada uno de los datos quedan de la siguiente forma:

$$\pi_1 = \frac{4 + 21}{50} = .5 \text{ y } \pi_2 = 1 - .5 = .5.$$

### 3.4 EJEMPLO CON TRES POBLACIONES

Vamos a repetir el anterior análisis suponiendo que tenemos tres poblaciones. Generamos 15 datos de una Normal(0,1), 15 de una Normal (0,2) y 15 de una Normal(1,.5), cada uno lo decidimos censurar generando 45 datos de una Normal(2,1), quedándonos con el mínimo entre las tres poblaciones y el dato censurado.

Población 1	Población 2	Población 3	Datos de Censura		
0.2087	-0.8525	2.5863	3.0562	3.4573	1.5217
-0.0978	0.0262	0.8543	0.5492	3.1033	1.0461
1.7871	-0.4201	1.5754	1.8747	1.0671	2.1973
1.6778	1.3661	1.4649	5.7220	4.2518	3.3995
-0.6102	-2.2019	2.3482	1.1711	0.9265	2.0699
0.3118	-0.2506	1.3399	2.1486	0.1115	2.3624
0.3763	1.1796	1.6331	3.0238	5.1193	1.4271
1.0907	-1.9276	0.5341	-0.4477	1.3539	4.7964
-0.5964	0.7628	1.4878	1.8242	0.4764	1.6266
0.2017	-0.2535	1.5422	4.4444	0.7977	1.1051
-0.3526	1.1568	0.0789	1.7604	7.9008	3.4516
-0.1370	0.8982	1.0071	0.5700	0.8923	6.0350
1.0529	0.5247	1.2637	4.4715	0.5683	3.7573
-0.5016	-0.0342	-0.1509	1.6057	2.5981	1.8851
-0.1338	0.6692	1.2713	1.7127	3.4081	1.2809

Para cada población obtenemos el mínimo entre el dato observado y el tiempo de censura, si el mínimo es el dato observado lo registramos en la población que pertenece, de lo contrario el dato lo registramos como censurado, sin saber a qué población pertenece.

Población 1	Población 2	Población 3	Censurado
0.2087	-0.8525	0.8543	-0.4477
-0.0978	0.0262	1.5754	0.4764
1.7871	-0.4201	1.4649	0.8923
1.6778	1.3661	1.3399	1.5217
-0.6102	-2.2019	0.5341	2.0699
0.3118	-0.2506	1.4878	1.4271

0.3763	1.1796	0.0789	1.1051
-0.5964	-1.9276	1.0071	1.2809
0.2017	-0.2535	1.2637	
-0.3526	1.1568	-0.1509	
-0.1370	0.5247		
1.0529	-0.0342		
-0.5016	0.6692		
-0.1338			

Tenemos 14 datos pertenecientes a la población 1, 13 datos pertenecientes a la población 2, 10 a la población 3 y 8 que no sabemos su verdadero valor ni a qué población pertenecen.

Para este ejemplo, debido a que el tamaño de la matriz de Fisher es de 8, no utilizaremos el método PLA, sino que directamente utilizaremos el método LVPLA. Aquí tenemos 3 poblaciones y 8 datos censurados, por lo que el número de posibilidades es de  $3^8=6561$ , este es el número total de casos que debemos contemplar para obtener la verosimilitud perfil.

Sin embargo, como estamos maximizando de forma independiente cada una de las tres poblaciones, en cada caso nos interesa saber únicamente, si el dato pertenece o no a la población, por lo que el número total de casos a maximizar es  $2^8=256$ , en cada población es decir en total tenemos que maximizar 768 posibilidades. Pero al momento de multiplicar las verosimilitudes perfil, sí tendremos que considerar los 6561 casos.

Al hacer esto tenemos los siguientes estimadores para los parámetros:

$$\hat{\mu}_1 = 0.29468, \quad \hat{\varphi}_1 = -0.28632, \quad \hat{\mu}_2 = 0.260475, \quad \hat{\varphi}_2 = 0.289845, \quad \hat{\mu}_3 = 1.1541537, \\ \hat{\varphi}_3 = -0.4663602.$$

Los estimadores de la varianza son:  $\sigma_1^2 = .56402$ ,  $\sigma_2^2 = 1.78548$  y  $\sigma_3^2 = .393481$ .

Si ahora calculamos las tres matrices observadas de Fisher tenemos:

$$I_1 = \begin{pmatrix} .03763 & .00255 \\ .00255 & .03531 \end{pmatrix}, I_2 = \begin{pmatrix} .12271 & .00531 \\ .00531 & .041233 \end{pmatrix} \text{ y } I_3 = \begin{pmatrix} .03263 & .00823 \\ .00823 & .05236 \end{pmatrix}$$

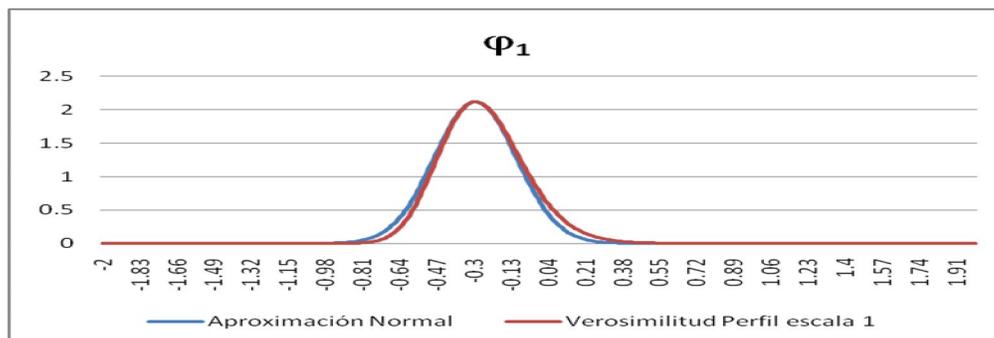
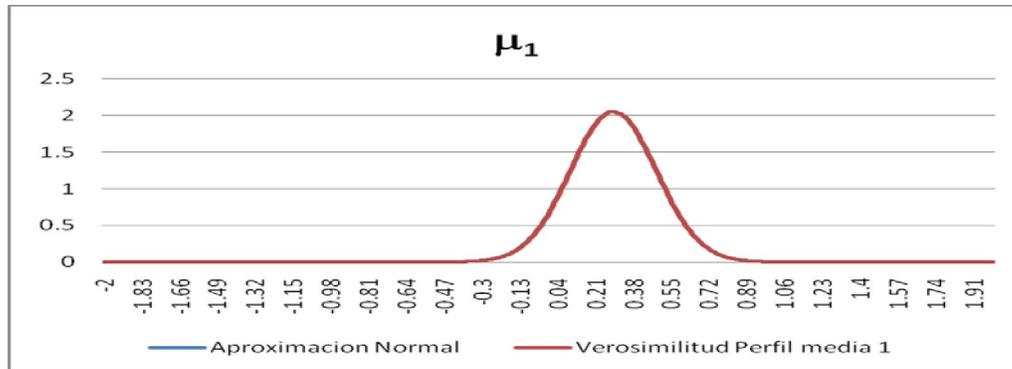
Con estos datos, los intervalos de confianza al 95% son:

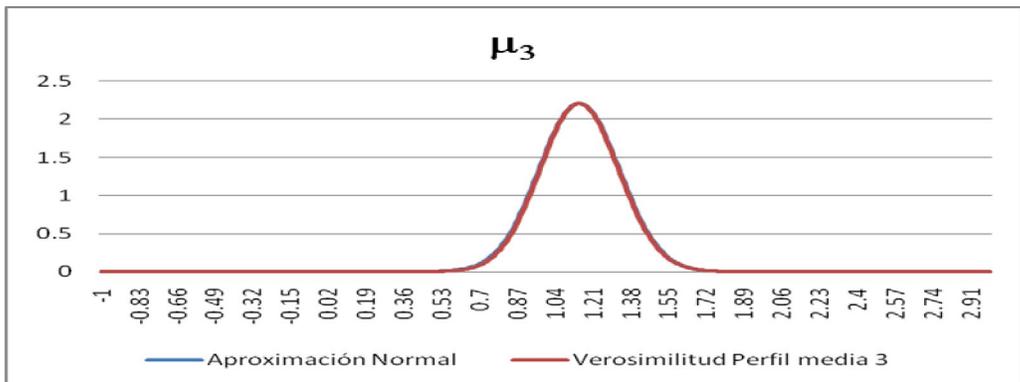
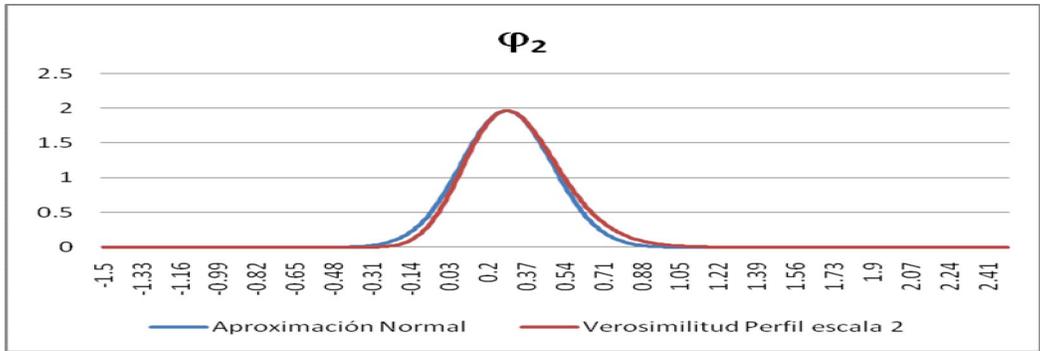
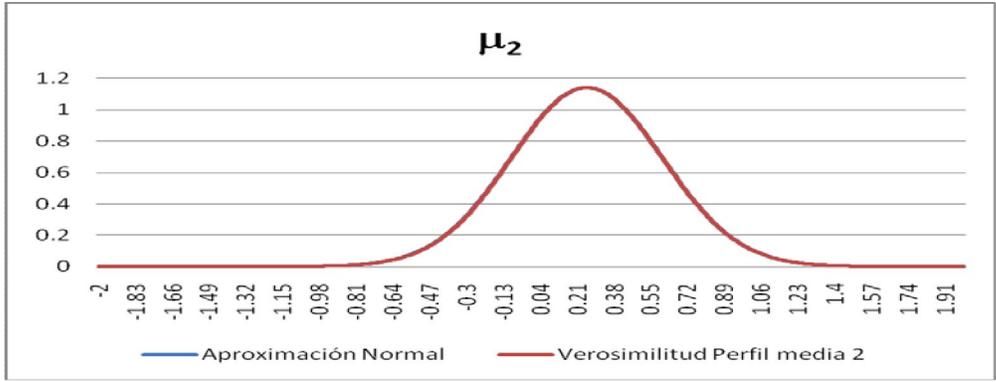
$$\mu_1 : (-.08556, .67493), \varphi_1 : (-.65462, .018967), \mu_2 : (-.42261, .94706), \varphi_2 : (-.108154, .68784), \\ \mu_3 : (.80006, 1.50824), \varphi_3 : (-.91486, -.017852).$$

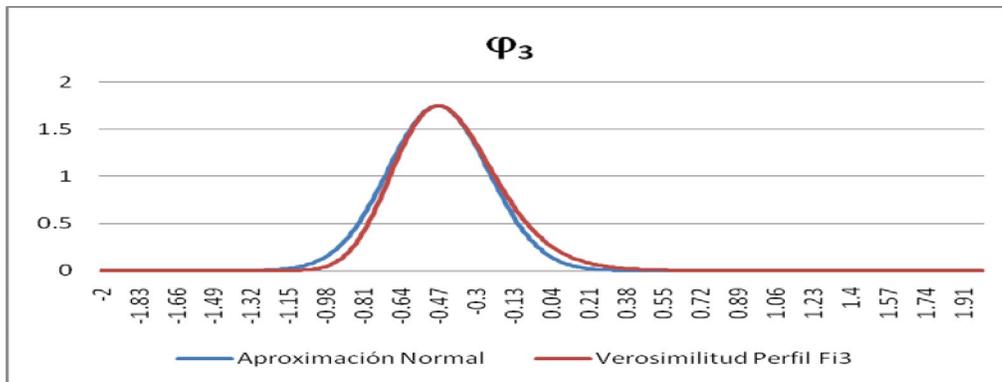
Para las varianzas los intervalos de confianza quedan de la siguiente manera:

$$\sigma_1^2 : (0.27002, 1.17817), \sigma_2^2 : (.805485, 3.957818) \text{ y } \sigma_3^2 : (.160455, .96492)$$

Las aproximaciones de las verosimilitudes perfil con la distribución normal con media el estimador máximo verosímil y varianza la matriz de Fisher correspondiente quedan como:







Como se ve en las anteriores gráficas la aproximación normal es muy buena con respecto a la verosimilitud perfil en cada uno de los casos.

La siguiente tabla muestra a donde fueron asignados los datos y a donde pertenecían originalmente.

Valor Censurado	Población Asignada	Verdadera Población
-0.4477	1	1
0.47641	1	2
0.8923	3	2
1.5217	2	3
2.0699	2	3
1.42709	3	3
1.1051	3	3
1.28091	3	3

Como se ve en la anterior gráfica, cuatro datos fueron asignados a una población a la que no pertenecían. Lo que hizo este método es asignar los valores más grandes a la población 2, que aunque su media es menor a la población tres, tiene la varianza más grande, mientras que la población 3 tiene la varianza más chica, por lo que puede tomar valores más grandes. Aunque algunos datos fueron agrupados de manera errónea, en realidad nos interesaba la estimación de los parámetros, la cual fue muy buena, ya que todos los intervalos de confianza contuvieron al verdadero valor del parámetro.

En realidad no nos interesa saber a qué población pertenece cada uno de los datos sino la proporción de datos de cada población, aunque este método nos permite asignar los datos censurados, al contrario del método PLA. Estimamos el número de datos

censurados en cada población:  $r_1=2$ ,  $r_2=2$  y  $r_3 =4$ . Con esto podemos estimar dichas proporciones:

$$\pi_1 = \frac{14+2}{45} = .3555555555, \pi_2 = \frac{13+2}{45} = .3333333333 \text{ y } \pi_3 = \frac{10+4}{45} = .3111111111$$

La proporción estimada de la segunda proporción es exactamente la verdadera proporción de la población 2. Para la proporción de la primera población podemos hacer una mezcla de Betas, con los siguientes parámetros: Beta(17,44) y Beta(16,45).

Si hacemos un intervalo de confianza al 95% para la proporción de la población 1, tenemos: (.1676,.3876), que contiene al verdadero valor de la proporción.

Para la proporción de la población 2, tenemos una mezcla de Betas con los siguientes parámetros: Beta(16,44) y Beta (15,45). Si hacemos el intervalo de confianza para la segunda proporción tenemos: (.1567,37521), el cual también contiene al verdadero valor de la segunda proporción.

Por último para la tercera población tenemos que la mezcla de Betas correspondientes es Beta(15,44) y Beta(14,45), por lo que el intervalo de confianza al 95% de la tercera población es: (.144559,.36223).

En este capítulo así como el anterior se utilizaron ejemplos generados artificialmente. En el siguiente capítulo se utilizará el método LVPLA para un caso citado en la literatura.

Lo que hace este método es separar los datos censurados en las diferentes poblaciones de acuerdo con el valor de cada uno de los datos, esto va a ser de vital importancia cuando se quiera simplificar el método LVPLA, como se verá más adelante.

#### **4. EJEMPLO**

En este capítulo se realiza una aplicación de la extensión del método LVPLA, desarrollado en el capítulo anterior, con un ejemplo de literatura.

El ejemplo con el cual se va a trabajar se tomó de Crowley J. y Hu, M (1977). El contexto del problema es el siguiente: tenemos un conjunto de pacientes a los cuales se les hizo un trasplante de corazón, y se siguió el tiempo de vida del paciente después de haber recibido dicho trasplante.

Si la muerte del paciente ocurrió antes del primero de abril de 1974, se registra el tiempo de vida, en días, del paciente después de recibir el trasplante, por otra parte si el paciente murió después del primero de abril de 1974, registramos solo el tiempo, en días transcurridos entre la fecha de corte y la fecha en la cual recibió el trasplante, en caso de que ocurriera lo último consideramos el dato como censurado.

Aparentemente nos encontramos en un caso de censura no aleatoria, ya que el tiempo de término de la observación está fijo. Sin embargo, la fecha en la cual el paciente recibió el trasplante varía de individuo a individuo, por lo que podemos modelar la censura como aleatoria.

Adicionalmente si el individuo murió antes del primero de abril de 1974, sabemos si éste murió a consecuencia de que su organismo rechazó el trasplante a corazón o por cualquier otra causa. En cambio para los individuos que murieron después del primero de abril de 1974, no sabemos ni el tiempo de su muerte, ni la causa de ésta.

De lo anterior podemos clasificar a los individuos cuya muerte fue registrada en dos poblaciones: 1) los que murieron a consecuencia del rechazo del trasplante, y 2) los que murieron por cualquier otra causa. Para los individuos censurados no podemos hacer lo anterior, por lo que usaremos el método LVPLA para clasificarlos. Tenemos la siguiente información:

-15 individuos de los cuales tenemos sus tiempos de muerte registrados y no rechazaron el trasplante.

-16 individuos de los que tenemos sus tiempos de muerte registrados y rechazaron el trasplante.

-10 individuos que fueron censurados y no tenemos sus tiempos de muerte, ni sabemos si rechazaron o no el trasplante.

Queremos ver si hay diferencia en el tiempo de supervivencia para los pacientes que rechazaron o no el trasplante.

En Crowley J. y Hu, M (1977), se analizó si existía diferencia o no entre los individuos que rechazaron el trasplante y los que no, construyendo la función de verosimilitud utilizando ciertas covariables, definidas en el artículo, a través de la función riesgo.

#### **4.1 EJEMPLO CON VALOR EXTREMO**

Como primer ejemplo, supondremos que el tiempo de muerte en días, tanto para las personas que no rechazaron el trasplante como las que si lo rechazaron es la densidad valor extremo, independientes entre sí.

Como vimos en el capítulo 1 (Gumbel), la función de densidad queda de la siguiente manera:

$$f(x, y) = \frac{e^{-\frac{(x-\mu_1)}{\sigma_1}} e^{-e^{-\frac{(x-\mu_1)}{\sigma_1}}}}{\sigma_1} \frac{e^{-\frac{(y-\mu_2)}{\sigma_2}} e^{-e^{-\frac{(y-\mu_2)}{\sigma_2}}}}{\sigma_2} \quad x \geq 0, y \geq 0$$

Los parámetros a estimar son  $\mu_1$ ,  $\sigma_1$ ,  $\mu_2$  y  $\sigma_2$ , para facilitar la aproximación a la normal, se utiliza las siguientes parametrizaciones:

$$\sigma_1 = e^{-\varphi_1} \quad \text{y} \quad \sigma_2 = e^{-\varphi_2} .$$

Haciendo esto, la función de densidad queda como:

$$f(x, y) = e^{(x-\mu_1)e^{-\varphi_1}} e^{-e^{(x-\mu_1)e^{-\varphi_1}}} e^{-\varphi_2} e^{(y-\mu_2)e^{-\varphi_2}} e^{-e^{(y-\mu_2)e^{-\varphi_2}}} e^{-\varphi_2} \quad x \geq 0, y \geq 0.$$

Sean  $x_{1j}$  los tiempos de muerte para los pacientes que no rechazaron el trasplante,  $x_{2j}$  los tiempos de muerte para los pacientes que rechazaron el trasplante,  $t_s$  con  $s=1, \dots, 11$ , los tiempos entre el trasplante y el 1 de abril de 1974, para los datos censurados. Adicionalmente, definimos  $\lambda_{is}$ , como una variable dicotómica, que toma el valor 1 si el  $s$ 'ésimo dato censurado se encuentra en la  $i$ 'ésima población. Las verosimilitudes perfiles son:

$$L_1(\mu_1, \varphi_1, x) = e^{-15\varphi_1} e^{\sum_{j=1}^{15} (x_{1j} - \mu_1)e^{-\varphi_1}} e^{-\sum_{j=1}^{15} e^{(x_{1j} - \mu_1)e^{-\varphi_1}}} \prod_{s=1}^{11} e^{-\lambda_{1s} e^{(t_s - \mu_1)e^{-\varphi_1}}}$$

$$L_2(\mu_2, \varphi_2, x) = e^{-15\varphi_2} e^{\sum_{j=1}^{15} (x_{2j} - \mu_2)e^{-\varphi_2}} e^{-\sum_{j=1}^{15} e^{(x_{2j} - \mu_2)e^{-\varphi_2}}} \prod_{s=1}^{11} e^{-\lambda_{2s} e^{(t_s - \mu_2)e^{-\varphi_2}}}$$

Las log-verosimilitudes quedan de la siguiente manera:

$$l_1(\mu_1, \varphi_1, x) = -15\varphi_1 + \sum_{j=1}^{15} (x_{1j} - \mu_1)e^{-\varphi_1} - \sum_{j=1}^{15} e^{(x_{1j} - \mu_1)e^{-\varphi_1}} - \sum_{s=1}^{11} \lambda_{1s} e^{(t_s - \mu_1)e^{-\varphi_1}}$$

$$l_2(\mu_2, \varphi_2, x) = -15\varphi_2 + \sum_{j=1}^{15} (x_{2j} - \mu_2)e^{-\varphi_2} - \sum_{j=1}^{15} e^{(x_{2j} - \mu_2)e^{-\varphi_2}} - \sum_{s=1}^{11} \lambda_{2s} e^{(t_s - \mu_2)e^{-\varphi_2}}$$

Tenemos que encontrar los estimadores máximos verosímiles para cada uno de los valores que pueden tomar  $\lambda_{ps}$ , que en este caso son:  $2^{10}=1024$ , para esto, construimos una matriz de tamaño 1024 por 10, que contiene unos y ceros que nos indican si pertenece o no a la población 1, mientras que el complemento nos servirá para la población 2. Una vez hecho esto, valuamos la función de verosimilitud, con las  $\lambda_{ps}$  y

los respectivos estimadores máximos verosímiles, y nos quedamos con la combinación que maximiza la verosimilitud. Los estimadores máximos verosímiles resultantes son:

$$\hat{\mu}_1 = 973.1433, \hat{\varphi}_1 = 6.7220, \hat{\mu}_2 = 515.3519 \text{ y } \hat{\varphi}_2 = 5.8402$$

Si aplicamos la parametrización para los parámetros de escala tenemos:

$$\hat{\sigma}_1 = 830.4764 \text{ y } \hat{\sigma}_2 = 543.8322.$$

La combinación de  $\lambda_{ps}$ , con que nos quedamos es:

Dato Censurado	Población Asignada
1775	1
1536	1
1549	1
660	1
955	1
305	2
339	2
30	2
237	2
167	2

Como se había mencionado en el capítulo anterior lo que hace este método es separar, en este caso los individuos censurados con mayor tiempo de vida los asignó a la población 1, que es la población que no rechazó el trasplante, mientras los individuos que vivieron menos fueron asignados a la población 2, los cuales si rechazaron el trasplante.

Si obtenemos las matrices observadas de Fisher, evaluadas en los estimadores máximos verosímiles y asignando cada dato censurado a una de las dos poblaciones de acuerdo con la tabla anterior tenemos:

$$I_1 = \begin{pmatrix} 46064.8933 & -1.8746 \\ -1.8746 & 0.0400 \end{pmatrix} \text{ y } I_2 = \begin{pmatrix} 7670.6672 & -2.8823 \\ -2.8823 & 0.0294 \end{pmatrix}$$

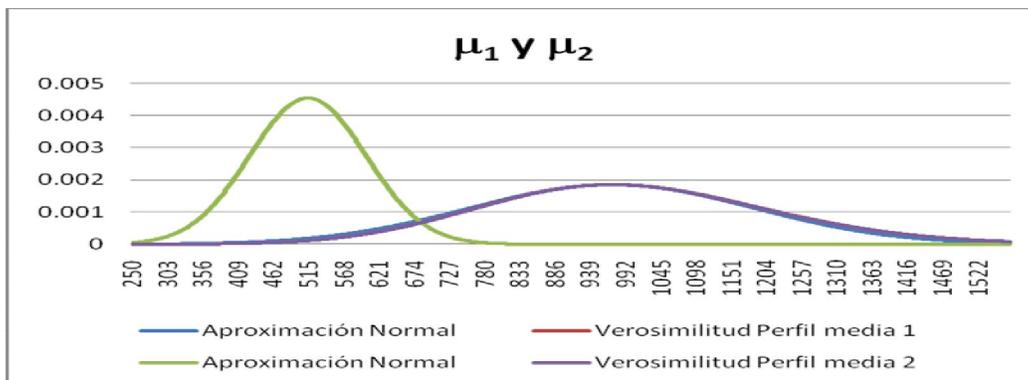
Construimos los intervalos de confianza al 95%, utilizando la aproximación normal, por lo que tenemos:

(552.4738, 1393.8129) para  $\mu_1$ , (6.3298, 7.1141) para  $\varphi_1$ , (343.6905, 687.0133) para  $\mu_2$  y (5.5042, 6.1760) para  $\varphi_2$

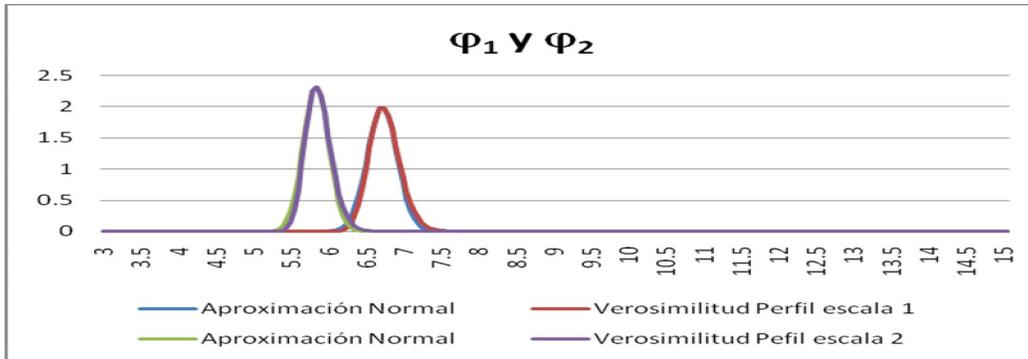
Para obtener los intervalos de confianza de los parámetros de escala aplicamos la transformación correspondiente para  $\varphi_1, \varphi_2$ :

(561.0951, 1229.1873) para  $\sigma_1$ , y (245.7409, 481.0779) para  $\sigma_2$ .

Ahora comparamos las verosimilitudes perfil de cada uno de los parámetros con respecto a su correspondiente aproximación normal, tomando como media el estimador máximo verosímil y como varianza la entrada correspondiente de la matriz observada de Fisher.



En el caso de los parámetros de localización se puede apreciar que la población 1, que no rechazó el trasplante, tiende a tener mayor tiempo de supervivencia que la población 2, en el cual si hubo rechazo. Aunque el parámetro de localización en el valor extremo no es exactamente la media, si nos sirve para conocer el valor aproximado que toman cada una de las poblaciones.



Para los parámetros de escala podemos ver que también se encuentran claramente separadas una población de otra. En este caso también la población 1 tiene un parámetro de escala mayor que la población 2.

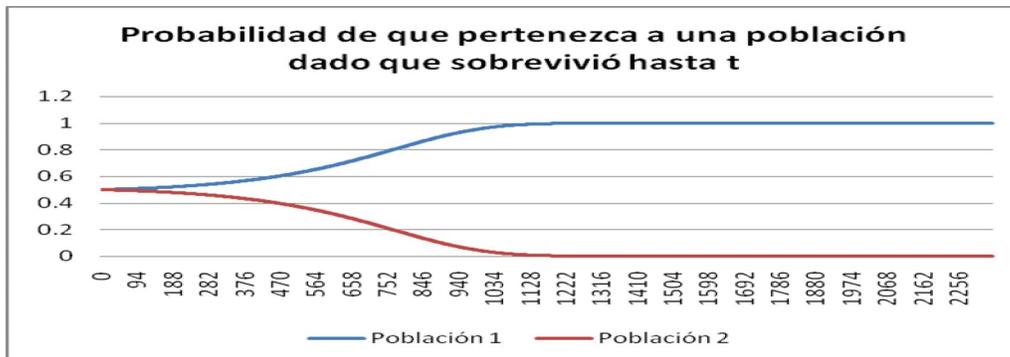
En conclusión, como se ve en ambas gráficas, las aproximaciones normales para las distribuciones perfil son muy buenas en los cuatro casos. Además de que en ambas gráficas se ve claramente que las poblaciones tienen parámetros muy diferentes entre sí.

Si ahora graficamos las funciones de supervivencia de cada población utilizando los parámetros estimados tenemos:



Como se puede ver una persona que no rechazó el trasplante tiene mayor probabilidad de sobrevivir que una persona que lo rechazó, utilizando las distribuciones de valor extremo.

Por último si decidimos graficar la probabilidad de que una persona sea de la población 1 ó 2, dado que sobrevivió hasta el tiempo  $t$ , con  $t > 0$ , tenemos:



Como se ve, en los resultados mostrados, este método sirvió para poder distinguir la supervivencia que tiene un paciente que rechazó el trasplante, con respecto a uno que no lo rechazó. Además, una vez con los estimadores máximos verosímiles de los parámetros de localización y escala podemos hacer predicciones sobre la supervivencia de los pacientes. Aquí podemos decir que sí hay diferencia estadística entre la supervivencia entre una población u otra, ya que si sabemos que sobrevivió a un tiempo  $t$ , grande, entonces es más probable que venga de la población 1.

La proporción de pacientes que rechazó el trasplante se estima como el número de pacientes que sabemos que no rechazaron el trasplante más el número asignado por el método dividido entre el número total de pacientes, es decir:

$$\hat{\pi}_1 = \frac{15 + 5}{41} = .4878,$$

Mientras que el número de personas que rechazaron el trasplante se obtiene por complemento de la anterior cantidad:

$$\pi_2 = 1 - .4878 = .5122$$

Si realizamos un intervalo de confianza al 95% para la proporción con una mezcla de Betas, las cuales son: Beta(21,21) y Beta(20,22) tenemos:

(.34001,.63724)

Lo que aporta este método, es que se puede aproximar las funciones de supervivencia para los individuos que no rechazaron el trasplante y los que sí, además de que se pueden hacer inferencias sobre los parámetros. En Crowley J. y Hu, M (1977) también se concluye que existe una diferencia entre el tiempo de vida después de trasplante de los individuos que rechazaron o no el trasplante, en el cual se hizo el estudio a través de covariables.

## 4.2 EJEMPLO CON DISTRIBUCIÓN NORMAL

El siguiente ejemplo que vamos a hacer es suponiendo que los tiempos de supervivencia después de haber recibido el trasplante se distribuyen normal para ambas poblaciones, y que también son independientes entre sí.

La función de densidad correspondiente es:

$$f(x_1, x_2) = \frac{1}{\sqrt{\sigma_1^{2n}}} e^{-\sum_{i=1}^{15} \frac{(x_{1i} - \mu_1)^2}{\sigma_1^2}} \frac{1}{\sqrt{\sigma_2^{2n}}} e^{-\sum_{i=1}^{15} \frac{(x_{2i} - \mu_2)^2}{\sigma_2^2}} .$$

Nuevamente utilizamos la anterior reparametrización para los parámetros de escala:

$$\sigma_1 = e^{-\varphi_1} \text{ y } \sigma_2 = e^{-\varphi_2} .$$

También en este caso tenemos:  $2^{10}=1024$  posibilidades, al asignar cada dato censurado a una de las dos poblaciones, mediante  $\lambda_{ps}$ . Por lo que las verosimilitudes perfil quedan como:

$$L_1(\mu_1, \varphi_1, x) = e^{-n\varphi_1} e^{-\sum_{i=1}^{15} \frac{1}{2}(x_{1i} - \mu_1)^2 e^{-2\varphi_1}} \prod_{s=1}^r (1 - \Phi((t_s - \mu_1)e^{-\varphi_1}))^{\lambda_{1s}}$$

$$L_2(\mu_2, \varphi_2, x) = e^{-n\varphi_2} e^{-\sum_{i=1}^{15} \frac{1}{2}(x_{2i} - \mu_2)^2 e^{-2\varphi_2}} \prod_{s=1}^r (1 - \Phi((t_s - \mu_2)e^{-\varphi_2}))^{\lambda_{2s}}$$

Las cuales las podemos trabajar por separado. Las log-verosimilitudes quedan como:

$$l_1(\mu_1, \varphi_1, x) = -n\varphi_1 - \sum_{i=1}^{15} \frac{1}{2}(x_{1i} - \mu_1)^2 e^{-2\varphi_1} + \sum_{s=1}^r \lambda_{1s} (1 - \Phi((t_s - \mu_1)e^{-\varphi_1}))$$

$$l_2(\mu_2, \varphi_2, x) = -n\varphi_2 - \sum_{i=1}^{15} \frac{1}{2}(x_{2i} - \mu_2)^2 e^{-2\varphi_2} + \sum_{s=1}^r \lambda_{2s} (1 - \Phi((t_s - \mu_2)e^{-\varphi_2})).$$

A continuación maximizamos las anteriores expresiones para todos las posibles combinaciones de  $\lambda_{ps}$ , obtenemos los estimadores máximo verosímiles, y valuamos en la función de verosimilitud y nos quedamos con la combinación que maximiza la función de verosimilitud. Los estimadores máximo verosímiles quedan como:

$$\hat{\mu}_1 = 575.6048, \quad \hat{\varphi}_1 = 6.663519, \quad \hat{\mu}_2 = 330.613, \quad \hat{\varphi}_2 = 5.7809.$$

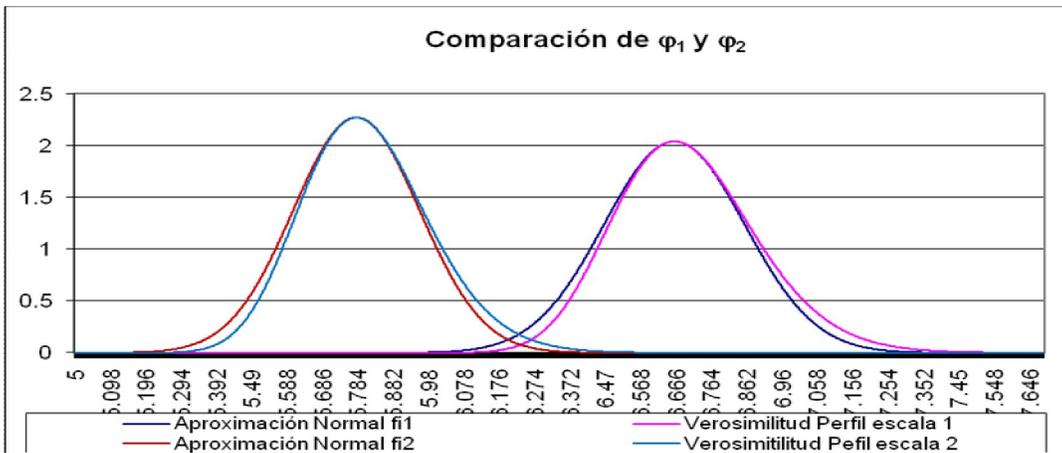
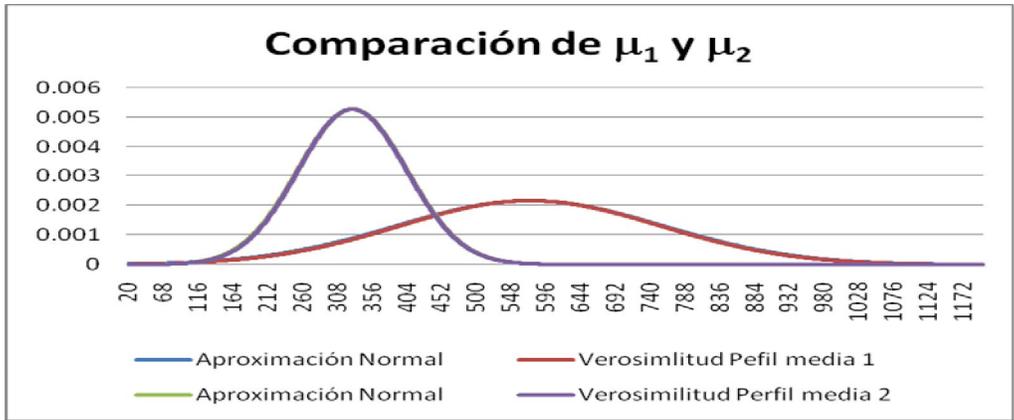
Las matrices observadas de Fisher quedan de la siguiente manera:

$$I_1 = \begin{pmatrix} 340005.36 & 5.86956 \\ 5.8956 & .038312 \end{pmatrix} \text{ y } I_2 = \begin{pmatrix} 5731.782 & 2.03337 \\ 2.03337 & .030764 \end{pmatrix}$$

Si hacemos los intervalos de confianza al 95% utilizando una aproximación normal tenemos los siguientes resultados:

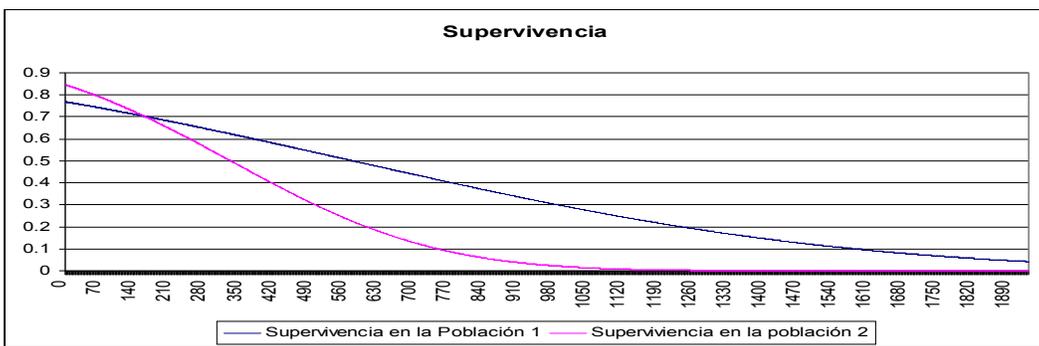
$$\mu_1 : (214.1701, 937.0934), \quad \varphi_1 : (6.279878, 7.04716), \quad \mu_2 : (182.2445, 479.002), \\ \varphi_2 : (5.437179, 6.124736).$$

A continuación comparamos las verosimilitudes perfil con sus respectivas aproximaciones normal:



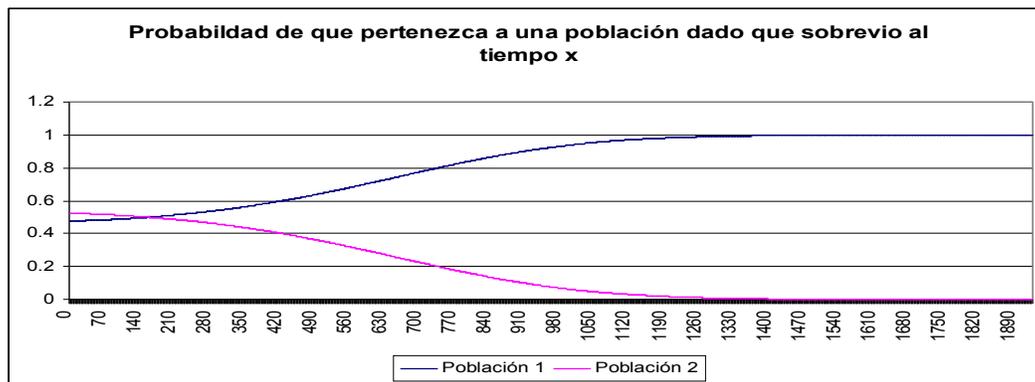
En ambas gráficas se puede ver que si hay diferencias tanto en los parámetros de localización como los de escala de población a población. Siendo que la población 1 tiene mayores valores en ambos parámetros.

Si graficamos la función de supervivencia para cada una de las poblaciones tenemos lo siguiente:



Como se puede ver en la gráfica anterior, para valores más pequeños las personas que rechazaron el trasplante tienen más probabilidad de sobrevivir que las personas que no rechazaron el trasplante, mientras que para más tiempo la probabilidad de sobrevivir es mayor en las personas que rechazaron el trasplante.

La siguiente gráfica describe la probabilidad de que una persona pertenezca a la población 1, dado que sobrevivió hasta el tiempo  $x$ , y la probabilidad de que pertenezca a la población 2 dado que sobrevivió hasta  $x$ .



Como se ve en la gráfica anterior, mientras aumenta el tiempo de supervivencia, aumenta la posibilidad de que pertenezca a la población 1, mientras que disminuye la posibilidad de que pertenezca a la población 2.

La combinación de  $\lambda_{ps}$ , que maximiza la verosimilitud es:

Dato Censurado	Población Asignada
1775	1
1536	1
1549	1
660	1
955	1
305	2
339	2
30	2
237	2
167	2

Como se puede apreciar, esta misma combinación era la que se obtiene si utilizamos la función valor extremo.

Además, en ambos casos los parámetros de localización y escala resultaron muy similares entre sí.

### **4.3 EJEMPLO CON DISTRIBUCIÓN WEIBULL**

Por último, el tercer ejemplo es manejar estos datos con la distribución Weibull: la función de densidad es:

$$f(x) = \alpha \theta x^{\alpha-1} e^{-\theta x^\alpha} \text{ si } x > 0.$$

Hacemos una reparametrización de  $\theta$ , con:  $\theta = e^\varphi$ , y tenemos:

$$f(x) = \alpha x^{\alpha-1} e^{-e^\varphi x^\alpha + \varphi}$$

Lo que tenemos que hacer es estimar los parámetros  $\varphi_1$ ,  $\alpha_1$ ,  $\varphi_2$ ,  $\alpha_2$ , mediante la asignación de los datos censurados. Como ya habíamos visto el número de posibilidades es 1024.

Las verosimilitudes perfil quedan de la siguiente manera:

$$L_i(\alpha_i, \varphi_i, x) = \prod_{j=1}^{n_i} \alpha_i x_{ij}^{\alpha_i-1} e^{-e^{\varphi_i} \sum_{j=1}^{n_i} x_{ij}^{\alpha_i} + n_i \varphi_i} \prod_{s=1}^r e^{-e^{\varphi_i} C_s^{\alpha_i} \lambda_{si}} \quad i=1,2.$$

La log-verosimilitudes que se obtienen son:

$$l_i(\alpha_i, \varphi_i, x) = n_i \ln(\alpha_i) + \sum_{j=1}^{n_i} (\alpha_i - 1) \ln(x_{ij}) - e^{\varphi_i} \sum_{j=1}^{n_i} x_{ij}^{\alpha_i} + n_i \varphi_i - e^{\varphi_i} \sum_{s=1}^r C_s^{\alpha_i} \lambda_{si} \text{ con } i=1,2.$$

Las derivadas parciales quedan como:

$$\frac{\partial l_i(\alpha_i, \varphi_i, \mathbf{x})}{\partial \alpha_i} = \frac{n_i}{\alpha_i} + \sum_{j=1}^{n_i} \ln(x_{ij}) - e^{\varphi_i} \sum_{j=1}^{n_i} \ln(x_{ij}) x_{ij}^{\alpha_i} - e^{\varphi_i} \sum_{s=1}^r \ln(C_s) C_s^{\alpha_i} \lambda_{si}$$

$$\frac{\partial l_i(\alpha_i, \varphi_i, \mathbf{x})}{\partial \varphi_i} = -e^{\varphi_i} \sum_{j=1}^{n_i} x_{ij}^{\alpha_i} + n_i - e^{\varphi_i} \sum_{s=1}^r C_s^{\alpha_i} \lambda_{si}$$

Tenemos 2 sistemas de ecuaciones de tamaño 2 que hay que resolver 1024 veces, de acuerdo con los distintos valores de  $\lambda_{si}$ . Al hacer esto tenemos que los estimadores máximo verosímiles son:

$$\hat{\alpha}_1 = .365198, \quad \hat{\varphi}_1 = -2.21038, \quad \hat{\alpha}_2 = .868838, \quad \hat{\varphi}_2 = -5.02781.$$

Si aplicamos la reparametrización para encontrar  $\lambda_i$ : tenemos.

$$\hat{\theta}_1 = .109659 \text{ y } \hat{\theta}_2 = .006553.$$

Las matrices observadas de Fisher, bajo esta distribución son:

$$I_1 = \begin{pmatrix} .006281 & -.03794 \\ -.03794 & .298569 \end{pmatrix}, \quad I_2 = \begin{pmatrix} .02775 & -.16621 \\ -.16221 & 1.057072 \end{pmatrix}$$

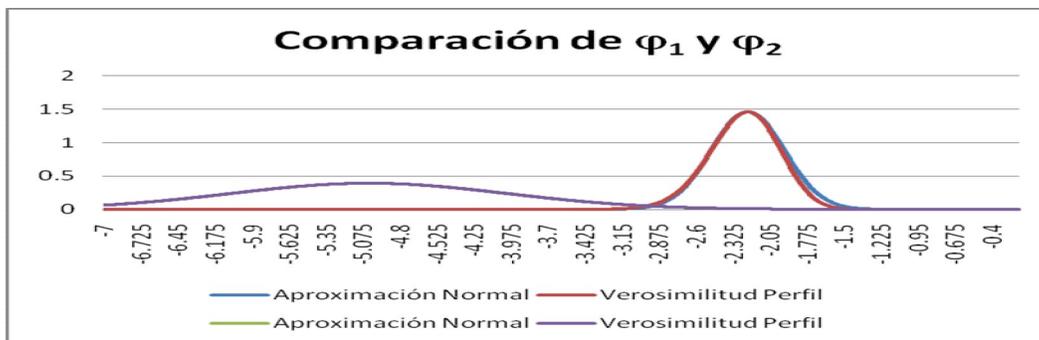
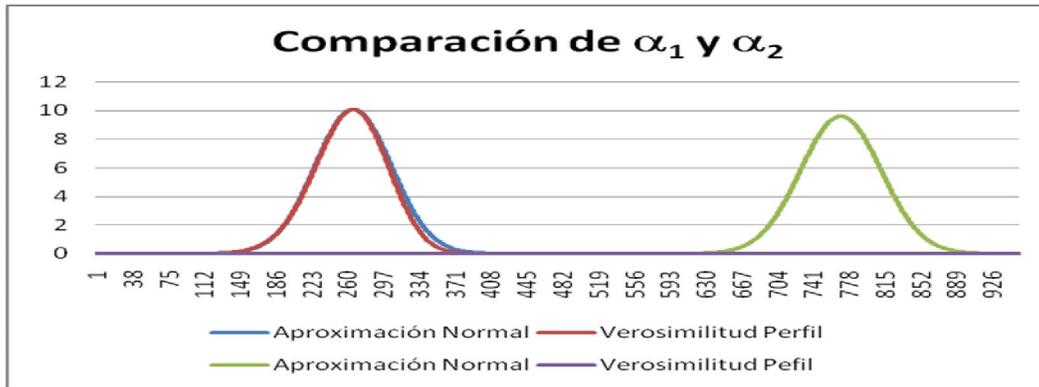
Con esta información calculamos los intervalos de confianza al 95% para todos los parámetros y obtenemos:

$$\alpha_1 : (.35288, .37509), \quad \varphi_1 : (-2.79029, -1.63048), \quad \alpha_2 : (.81439, .923277), \\ \varphi_2 : (-7.99067, -2.9559).$$

Para las  $\theta_i$ 's los intervalos quedan como:

$\theta_1 : (.06141, .195836)$  y  $\theta_2 : (.000825, .052029)$ .

Por último, graficamos las funciones de verosimilitudes perfil con las respectivas aproximaciones normal tenemos:

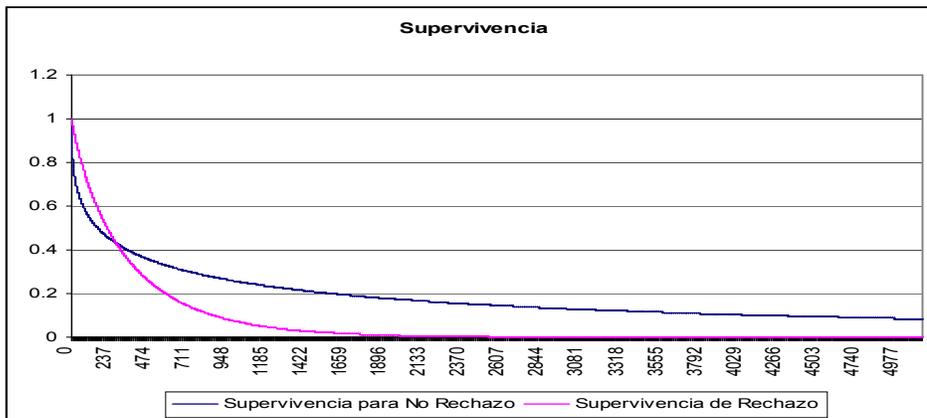


Las aproximaciones normales son muy buenas para las verosimilitudes perfil, además se puede ver que existe mucha diferencia entre ambos parámetros de la población.

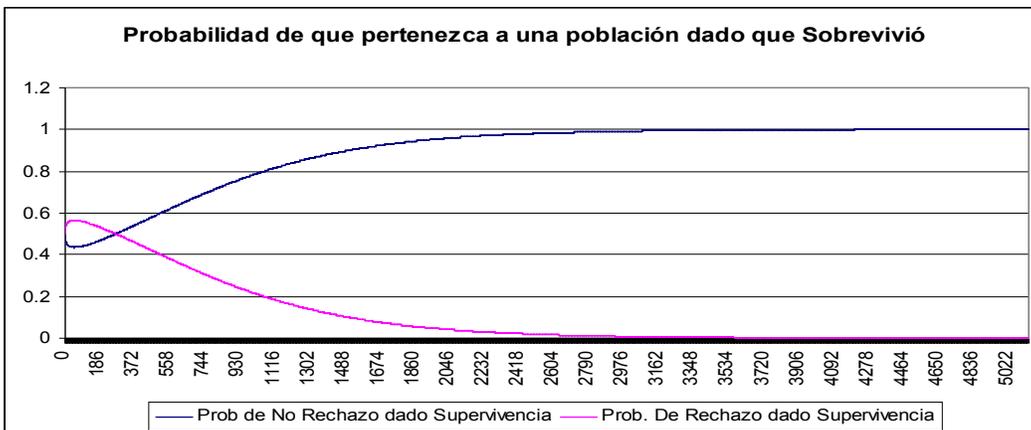
Dato Censurado	Población Asignada
1775	1
1536	1
1549	1
660	1
955	1
305	2
339	2
30	2
237	2
167	2

Como se puede apreciar la asignación fue la misma que en los dos anteriores casos, esto es debido a que los datos están muy separados entre sí, por lo que podemos hablar de dos grupos.

Si graficamos las funciones de supervivencia con los dos juegos de parámetros tenemos:



Como se ve en la anterior gráfica para valores pequeños es más probable que sobreviva una persona que rechazó el trasplante, pero para valores más altos hay mayor probabilidad en los que no rechazaron el trasplante.



La gráfica anterior muestra la probabilidad de que una persona haya rechazado el trasplante o no dado que sabemos que sobrevivió al tiempo t. Como se puede ver para

valores muy grandes de  $t$ , la probabilidad de que no haya rechazado el trasplante es prácticamente de 1, mientras que la probabilidad de que lo haya rechazado es 0.

En las tres gráficas donde comparábamos la supervivencia y la pertenencia a una población dado que sobrevivió a un tiempo  $t$ , presentan aspectos similares. En la gráfica de supervivencia, se ve que la función de supervivencia cae más rápido en la segunda población que en la primera para las tres distribuciones. Para la gráfica de pertenencia a una población, para la población 1 la probabilidad tiende a uno para valores más grandes, por lo que la probabilidad de pertenecer a la población 2 tiende a cero.

Aunque en los tres ejemplos se supuso que las dos poblaciones tenían la misma función de densidad, esto no es necesariamente cierto, pudimos haber supuesto una distribución normal para una población y densidad de valor extremo para la otra. Los datos de donde se obtuvieron dichos resultados vienen en el anexo, al final de este trabajo.

## **5. SIMPLIFICACIÓN DEL MÉTODO LVPLA BAJO CENSURA ALEATORIA**

En el problema que estamos considerando, donde suponemos que tenemos  $r$  datos censurados que pueden pertenecer a " $k$ " poblaciones, vimos que hay que realizar, bajo el método LVPLA,  $2^r$  optimizaciones para cada una de las " $k$ " poblaciones, ya que tenemos que considerar para cada dato censurado si pertenece o no a la población en cuestión, por lo que hay que realizar  $k2^r$  optimizaciones, el cual puede ser un número muy grande. Una vez hechas este número de optimizaciones, hay que considerar todas las combinaciones posibles, para los  $r$  datos censurados en cada una de las " $k$ " poblaciones, es decir,  $k^r$ , ya que cada dato tiene " $k$ " posibilidades al pertenecer a una y solamente una de las poblaciones, el cual puede ser un número más grande todavía que el anterior.

Como se ha visto en los ejemplos desarrollados en los anteriores capítulos, una de las propiedades del método LVPLA, consiste en separar los datos censurados de acuerdo con la magnitud de éstos. El objetivo de este capítulo consiste en primero probar que el método LVLPA siempre cumple con esta propiedad. Después de haber hecho esto, aprovecharemos dicha propiedad para mejorar el método.

### **5.1 DEMOSTRACIÓN DE LA SEPARACIÓN DE DATOS DE ACUERDO A SU MAGNITUD POR EL MÉTODO LVPLA**

#### **5.1.1 Teorema:**

Supongamos que tenemos un conjunto de datos pertenecientes a " $k$ " poblaciones bajo censura aleatoria, por lo que tenemos dos situaciones:

- 1) Si el tiempo de falla ocurrió antes que la censura, se registra dicho tiempo y se clasifica a la población que pertenece, estos valores lo denotamos por  $x_{ij}$ ,  
 $i=1 \dots k, j=1 \dots n_i$

- 2) Si el tiempo de falla ocurrió después del tiempo de censura, no registramos el valor, ni la población a la que pertenece. Solo sabemos la existencia de dicho valor y el tiempo de censura, el cuál en el caso de la censura por la derecha, nos indica que el tiempo donde habría ocurrido la falla sería mayor, estos valores se denotan por  $t_s, s=1, \dots, r$

Es decir, estamos suponiendo que tenemos  $r$  datos donde la censura ocurrió antes, es decir, que no sabemos su verdadero valor, ni la población a la que pertenece. Supongamos, además, que  $r$  y  $k$  son tales que  $k2^r$  y  $k'$  son números muy grandes, por lo que no sería factible hacer una enumeración de todos los casos.

Por el método LVPLA, encontramos los estimadores máximo verosímiles, para  $\mu, \sigma$ , así como la asignación de las  $t_s, s=1, \dots, r$ , en cada una de las poblaciones.

Entonces, si una  $t_p$  en particular fue asignada a una población  $c$ , en el máximo, existe una vecindad con centro en  $t_p$ , que también será asignada a la población  $c$  y que también maximizará la función de verosimilitud.

### Demostración:

La función de verosimilitud con censura aleatoria es:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{x}) = \frac{r}{r_1 r_2 \dots r_k} \prod_{i=1}^k \prod_{s=1}^r (1 - F_i(t_s, \mu_i, \sigma_i))^{\lambda_{si}} \prod_{i=1}^k \prod_{j=1}^{n_i} f_i(x_{ij}, \mu_i, \sigma_i).$$

Supongamos que por el método LVPLA se encuentran  $\mu_i^*, \sigma_i^*, \lambda_{si}^*$ , tales que la verosimilitud es óptima:

$$L^*(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*, \mathbf{x}) = \frac{r}{r_1 r_2 \dots r_k} \prod_{i=1}^k \prod_{s=1}^r (1 - F_i(t_s, \mu_i^*, \sigma_i^*))^{\lambda_{si}^*} \prod_{i=1}^k \prod_{j=1}^{n_i} f_i(x_{ij}, \mu_i^*, \sigma_i^*).$$

Supongamos que tomamos una de las  $t_s$ , llamada  $t_p$ , y esta se asigna a la población  $c$ , en el máximo, es decir,  $\lambda_{pc}=1$ .

Entonces como  $t_p$ , se tiene que:

$$\frac{r!}{r_1!r_2!\dots r_c!r_k!}(1 - F_c(\mu_c, \sigma_c, t_x)) > \frac{r!}{r_1!r_2!\dots(r_c - 1)!(r_j + 1)!\dots r_k!}(1 - F_j(\mu_j, \sigma_j, t_x))$$

$j \neq c$ , ya que de lo contrario sería asignado a otra población. Como vale para toda  $j \neq c$ , entonces se cumple:

$$\frac{r!}{r_1!r_2!\dots r_c!r_k!}(1 - F_c(t_p, \mu_c, \sigma_c)) > \max_{j \neq c} \left( \frac{r!}{r_1!r_2!\dots(r_c - 1)!(r_j + 1)!\dots r_k!}(1 - F_j(\mu_j, \sigma_j, t_p)) \right).$$

Sabemos que:  $\frac{r!}{r_1!r_2!\dots r_c!r_k!}(1 - F_c(\mu_c, \sigma_c, t_p))$  es una función continua con respecto a  $t_p$

y  $\frac{r!}{r_1!r_2!\dots(r_c - 1)!(r_j + 1)!\dots r_k!}(1 - F_j(\mu_j, \sigma_j, t_p))$  también es una función continua con

respecto a  $t_p$  para toda  $j \neq c$ , por lo que:

$\max_{j \neq c} \left( \frac{r!}{r_1!r_2!\dots(r_c - 1)!(r_j + 1)!\dots r_k!}(1 - F_j(\mu_j, \sigma_j, t_p)) \right)$  también es una función continua

con respecto a  $t_p$ .

Definimos:

$$g(t_x) = \frac{r!}{r_1!r_2!\dots r_c!r_k!}(1 - F_c(\mu_c, \sigma_c, t_x)) - \max_{j \neq c} \left( \frac{r!}{r_1!r_2!\dots(r_c - 1)!(r_j + 1)!\dots r_k!}(1 - F_j(\mu_j, \sigma_j, t_x)) \right).$$

la cual también es una función continua con respecto a  $t_x$ , y además  $g(t_p) > 0$ .

Por lo tanto existe una  $\delta > 0$ , tal que si  $|t_x - t_p| < \delta$  entonces  $g(t_x) > 0$

La anterior demostración utiliza básicamente la definición de continuidad y límites.

$g(t_x) > 0$  en una vecindad de  $t_p$ , entonces en dicha vecindad se cumple:

$$\frac{r!}{r_1!r_2!\dots r_c!r_k!}(1 - F_c(\mu_c, \sigma_c, t_x)) > \max\left(\frac{r!}{r_1!r_2!\dots(r_c - 1)!(r_j + 1)!\dots r_k!}(1 - F_j(\mu_j, \sigma_j, t_x))\right)$$

Por lo que concluimos que si  $|t_x - t_p| < \delta$ , entonces  $t_x$  sigue siendo asignado a la población  $c$ , en el máximo.

## **5.2 PASOS PARA SIMPLIFICAR EL MÉTODO LVPLA**

Supongamos que nos encontramos bajo el mismo contexto que el problema anterior, es decir, tenemos un conjunto de datos pertenecientes a “ $k$ ” poblaciones, bajo censura aleatoria, entonces podemos encontrarnos con dos situaciones.

- 1) Si el tiempo de falla ocurrió antes que la censura, se registra dicho tiempo y se clasifica a la población que pertenece.
- 2) Si el tiempo de falla ocurriría después del tiempo de censura, no registramos el valor, ni la población a la que pertenece. Solo sabemos la existencia de dicho valor y el tiempo de censura, el cuál en el caso de la censura por la derecha, nos indica que el tiempo donde habría ocurrido la falla sería mayor.

Supongamos que tenemos  $r$  datos donde la censura ocurrió antes, es decir, que no sabemos su verdadero valor, ni la población a la que pertenece. Supongamos, además, que  $r$  y  $k$  son tales que  $k2^r$  y  $k^r$  son números muy grandes, por lo que no sería factible tomar todos los casos.

### **5.2.1 PASO 1**

El primer paso que se haría es generar  $k$  matrices de tamaño igual al número de escenarios que se deseen por  $k$ , cuyas entradas contienen uno cero, es decir, que el escenario es uno de los posibles casos donde asignamos un dato censurado a una y solamente una de las poblaciones.

Como cada renglón de cada una de dichas matrices nos indica si un dato censurado está o no en la población entonces estas matrices deben cumplir con la propiedad de que las  $\lambda_{is}$  sumen uno. En este caso, no estamos abarcando todas las posibilidades, sino solo un subconjunto de éstas.

Con estas “ $k$ ” matrices a las que denotaremos como  $A_p$ , cuya entrada  $a_{ij}$ , nos indica si el  $i$ 'ésimo dato censurado pertenece a la  $p$ -ésima población, en el escenario  $j$ , usamos el método LVPLA, pero no para todos los  $k^r$  casos, sino para los escenarios que estamos considerando.

Al hacer esto, obtenemos un vector de verosimilitudes de tamaño igual número de escenarios. Decidimos quedarnos, únicamente con el 10% de los valores más grandes de dicho vector, el cual llamaremos,  $L$ , así como sus correspondientes 10% de los renglones de las matrices  $A_p$ , las cuales denotaremos como  $B_p$ , que son matrices de tamaño del 10% del número de escenarios por  $k$ .

Para cada una de estas matrices, calculamos la siguiente ponderación:

$$P_{is} = \frac{B'_i(:,s) * L}{U * L},$$

donde  $U$  es un vector de tamaño 1 por el 10% del número de escenarios, que tiene únicamente unos en sus entradas, y al multiplicarlo por  $L$ , está haciendo que se sumen todas las entradas de  $L$ .

La anterior cantidad representa una ponderación de las columnas de  $B$ , con respecto a la verosimilitud, como  $B(:,s)$  es una columna que tiene únicamente unos y ceros, solo va a sumar los que fueron asignados a esa población, por lo que al dividirlo entre la suma de las verosimilitudes tenemos una especie de “probabilidad” de que el  $s$ ’ésimo dato censurado pertenezca a la  $i$ ’ésima población.

Para terminar este paso, se analiza cuales datos censurados tienen “probabilidades” grandes en cada población y cuales tienen valores chicos, es decir, queremos ver de que manera separa cada uno de los datos censurados a una población de acuerdo a su magnitud.

### **5.2.2 PASO 2**

El segundo paso consiste en que, una vez visto las regiones que se generaron, con respecto al paso anterior, se vayan variando los márgenes, de dichas regiones y llevándose a cabo el método LVPLA. El número de casos es considerablemente más chico.

### **5.3. EJEMPLO CON DOS POBLACIONES**

Generamos dos poblaciones Normales con 25 datos cada una, la primera es Normal(0,1), mientras que la segunda es: Normal(1,1.5). Ambas se censuran aleatoriamente por la derecha con dos Normales (2,2). Los datos se presentan a continuación:

Normal (0,1)	Normal(1,1.5)	Normal (2,2)	
2.4594	0.5539	-0.3692	1.4895
1.0482	1.9396	-0.8390	3.2404
1.7076	3.1355	-1.2465	2.4460

1.0795	1.8021	4.4904	1.0898
-0.7614	0.5719	2.0183	0.8473
-0.6409	2.5334	1.3608	0.6467
-0.8017	1.1036	2.5116	0.7736
-0.4417	1.2359	1.4082	3.8374
-1.3586	-0.7826	0.9736	0.7262
1.3766	1.9628	3.4391	1.2534
-0.4858	1.7876	2.9104	1.8118
-0.2470	0.3633	0.9389	3.4517
2.2835	0.4451	-0.6041	0.9542
-1.4856	2.5047	3.6169	3.0369
0.3015	-2.0540	2.9669	2.1437
-0.9076	1.7096	-0.1850	2.4780
0.2741	0.1879	0.6479	1.2590
-0.7924	0.6064	2.6550	1.7484
0.1067	1.8311	4.5256	1.6273
-0.7647	1.7314	2.9976	0.7660
-0.3247	2.3235	2.1796	3.3839
-0.5334	1.8106	0.9209	2.7833
-1.5287	2.7300	1.7976	2.1276
0.3895	0.5370	2.2467	0.9705
-0.7907	0.0553	1.6382	0.9648

Al quedarnos con el mínimo tenemos lo siguiente:

Población 1	Población 2	Censurados
1.0795	0.5539	-0.3692
-0.7614	1.9396	-0.8390
-0.6409	0.5719	-1.2465
-0.8017	1.2359	-0.6041
-0.4417	-0.7826	2.4460
-1.3586	1.7876	1.0898
1.3766	0.3633	0.6467
-0.4858	0.4451	0.7736
-0.2470	2.5047	1.2534
-1.4856	-2.0540	1.6273
0.3015	1.7096	0.7660
-0.9076	0.1879	2.1276
0.2741	0.6064	
-0.7924	2.3235	

0.1067	1.8106	
-0.7647	0.5370	
-0.3247	0.0553	
-0.5334		
1.5287		
0.3895		
-0.7907		

Tenemos 21 datos que sabemos pertenecen a la población 1, 17 a la población 2, y 12 datos censurados.

Si lo hiciéramos por el método LVPLA, habría que considerar  $2^{12}=4096$  casos, por lo que decidimos probarlo con la simplificación:

Generamos una matriz,  $A_1$ , de tamaño 1000 por 12, con una Bernouilli con parámetro .5, mientras  $A_2=1-A_1$ . Hacemos el método LVPLA, para los 1000 casos, obtenemos los 1000 valores de la verosimilitud perfil, nos quedamos únicamente con los 100 valores más grandes, tanto de la verosimilitud, como de los renglones correspondientes a  $A_1$  y  $A_2$ , y encontramos los valores de las  $p_{is}$ :

Datos	Prob. 1	Prob 2
-0.3692	0.5242	0.4758
-0.8390	0.7417	0.2583
-1.2465	0.6343	0.3657
-0.6041	0.5872	0.4128
2.4460	0	1
1.0898	0.2445	0.7555
0.6467	0.4304	0.5696
0.7736	0.2021	0.7979
1.2534	0.1743	0.8257
1.6273	0.0758	0.9242
0.7660	0.2357	0.7643
2.1276	0.0269	0.9731

Como se ve en la gráfica, los valores pequeños tienen probabilidad alta de estar en la población 1, mientras que los valores grandes tienen probabilidad alta de estar en la población 2. Para hacer todo esto el programa tardó 50 segundos.

Entonces vimos que lo que hace LVPLA es separar en dos regiones los valores pequeños van a la población 1, mientras que los valores grandes a la población 2, por lo que podemos hablar de una barrera que separa a estas dos regiones. El paso dos consistiría en ver donde se encuentra esta barrera.

Consideramos ahora 13 casos: el primero sería que ningún valor se asigna a la población 1, el segundo que solo el más pequeño, el tercero que los dos más pequeños a la población 1 y el resto a la dos, y así, hasta el último caso donde todos se asignan a la población 1. Hacemos el LVPLA con estos casos y obtenemos los estimadores máximo verosímiles:

$$\hat{\mu}_1 = -0.18781, \hat{\varphi}_1 = -0.235096, \hat{\mu}_2 = 1.31759, \hat{\varphi}_2 = 0.254516.$$

Para calcular esto tardó .56 segundos. Las matrices observadas de Fisher son:

$$I_1 = \begin{pmatrix} 0.02764 & 0.00223214 \\ 0.00223214 & 0.0229640 \end{pmatrix} \text{ y } I_2 = \begin{pmatrix} 0.07964963 & 0.0112895 \\ 0.0112895 & 0.0310816 \end{pmatrix}$$

Los intervalos de confianza al 95% son:

$$\mu_1 : (-0.51368, 0.13806) \quad \varphi_1 : (-0.53211, 0.06192) \quad \mu_2 : (0.7644378, 1.871) \text{ y} \\ \varphi_2 : (-.0910316, .6000637)$$

Los intervalos para las varianzas son:

$$\sigma_1^2 : (.334499, 1.131886) \text{ y } \sigma_2^2 : (.835487, 1.3.3205439)$$

Todos los intervalos contienen el verdadero valor del parámetro.

Para confirmar la eficiencia de la simplificación se hizo el método LVPLA para los 4096 casos y se llegó a la misma asignación de los datos censurados a cada una de las poblaciones y por lo tanto a los mismos estimadores máximo verosímiles.

Datos	Población Asignada	Población Verdadera
-0.3692	1	1
-0.8390	1	1
-1.2465	1	1
-0.6041	1	1
2.4460	2	2
1.0898	2	2
0.6467	2	2
0.7736	2	2
1.2534	2	2
1.6273	2	2
0.7660	2	2
2.1276	2	2

Como se puede ver en la anterior tabla, todos los datos censurados fueron asignados correctamente, como ya habíamos mencionado, esto no siempre ocurre, pero lo que nos interesaba demostrar es que con la simplificación y el método LVPLA se llega a lo mismo. El método LVPLA completa tarda en este caso 208 segundos.

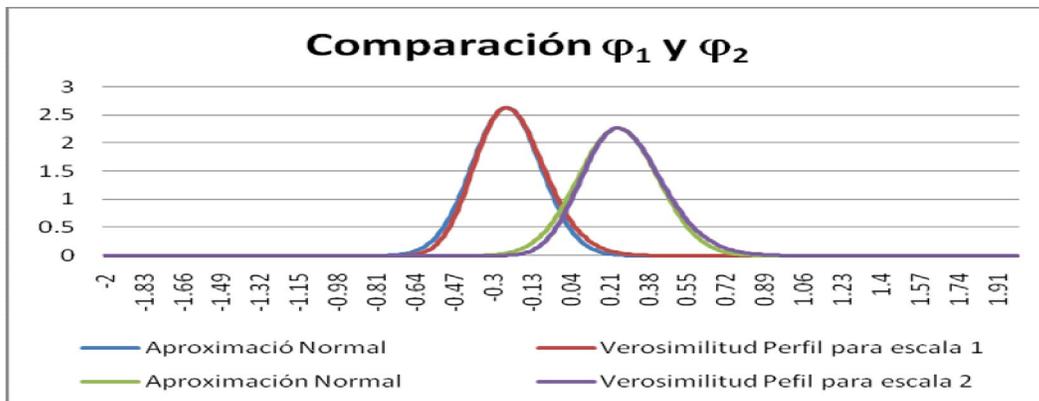
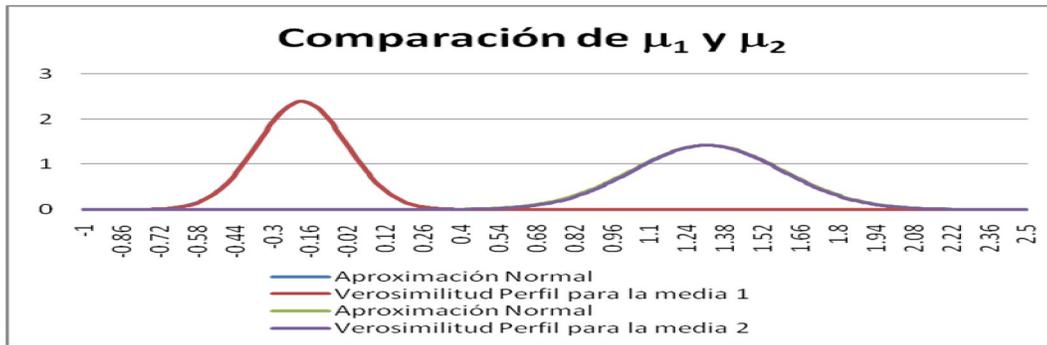
La siguiente tabla muestra las verosimilitudes correspondientes al ir moviendo el margen:

Margen	Verosimilitud
$-\infty$	3.40E-15
-1.2465	2.42E-10
-0.8390	1.18E-09
-0.6041	3.16E-09
-0.3692	5.11E-09
0.6467	2.10E-09
0.7660	6.52E-10
0.7736	1.80E-10
1.0898	3.12E-11
1.2534	4.37E-12
1.6273	4.33E-13
2.1276	3.47E-14
2.4460	3.40E-15

El margen es el máximo valor que se asigna a la población 1. En caso de que sea  $-\infty$ , significa que ninguno fue asignado a la población 1, y como se puede ver, la

verosimilitud alcanza su máximo cuando el máximo valor que se asigna a la población 1 es  $-0.3692$ .

Por último se comparan las verosimilitudes perfil con sus correspondientes aproximaciones normales.



De las anteriores gráficas, se observa que la aproximación normal es muy buena con respecto a la verosimilitud perfil. Además las medias presentan gran diferencia, al contrario de los parámetros de escala que son muy parecidos.

La siguiente tabla contiene el valor de la verosimilitud, únicamente contando los datos censurados, si al valor que consideramos como frontera,  $-0.3692$ , lo asignamos a la población 1, y si lo asignamos a la población 2:

	Población Asignada	Verosimilitud	Población Asignada	Verosimilitud
--	--------------------	---------------	--------------------	---------------

-1.2465	1	0.9098	1	0.9098
-0.839	1	0.7950	1	0.7950
-0.6041	1	0.7008	1	0.7008
<b>-0.3692</b>	<b>1</b>	<b>0.5907</b>	<b>2</b>	<b>0.9045</b>
0.6467	2	0.6985	2	0.6985
0.766	2	0.6655	2	0.6655
0.7736	2	0.6634	2	0.6634
1.0898	2	0.5701	2	0.5701
1.2534	2	0.5198	2	0.5198
1.6273	2	0.4051	2	0.4051
2.1276	2	0.2650	2	0.2650
2.446	2	0.1908	2	0.1908
Verosimilitud		<b>0.5612</b>		<b>0.3125</b>

Como se puede ver el dato  $-0.3692$ , si toma el máximo si lo asignamos a la población 1. La siguiente tabla muestra el valor de la frontera tal que hace que las verosimilitudes sean iguales:

	Población Asignada	Verosimilitud	Población Asignada	Verosimilitud
-1.2465	1	0.9098	1	0.9098
-0.839	1	0.7950	1	0.7950
-0.6041	1	0.7008	1	0.7008
<b>0.2510</b>	<b>1</b>	<b>0.2894</b>	<b>2</b>	<b>0.7959</b>
0.6467	2	0.6985	2	0.6985
0.766	2	0.6655	2	0.6655
0.7736	2	0.6634	2	0.6634
1.0898	2	0.5701	2	0.5701
1.2534	2	0.5198	2	0.5198
1.6273	2	0.4051	2	0.4051
2.1276	2	0.2650	2	0.2650
2.446	2	0.1908	2	0.1908
Verosimilitud		<b>0.2749</b>		<b>0.2749</b>

Entonces la vecindad para el dato de la frontera que hace que las verosimilitudes sean iguales es  $.2510$ , por lo que la vecindad donde el valor de la frontera sigue siendo asignado al máximo es  $(-\infty, .2510)$ .

#### **5.4. EJEMPLO CON TRES POBLACIONES**

Este ejemplo se tomo de los datos del capítulo 3, cuando se utilizó el método LVPLA con tres poblaciones. En aquel caso habíamos generados tres poblaciones normales, cada una con 15 datos: Normal(0,1), Normal (0,2) y Normal(1,.5), y se censuran con una Normal(2,1), quedándonos con el mínimo entre las tres poblaciones y el dato censurado. Los datos con los que contábamos eran los siguientes:

Población 1	Población 2	Población 3	Censurado
0.2087	-0.8525	0.8543	-0.4477
-0.0978	0.0262	1.5754	0.4764
1.7871	-0.4201	1.4649	0.8923
1.6778	1.3661	1.3399	1.5217
-0.6102	-2.2019	0.5341	2.0699
0.3118	-0.2506	1.4878	1.4271
0.3763	1.1796	0.0789	1.1051
-0.5964	-1.9276	1.0071	1.2809
0.2017	-0.2535	1.2637	
-0.3526	1.1568	-0.1509	
-0.1370	0.5247		
1.0529	-0.0342		
-0.5016	0.6692		
-0.1338			

Tenemos 14 datos pertenecientes a la población 1, 13 datos pertenecientes a la población 2, 10 a la población 3 y 8 que no sabemos su verdadero valor ni a qué población pertenecen.

El número de optimizaciones que tenemos que hacer es  $3(2)^8=768$ , y para construir todos los casos de la verosimilitud tenemos que tomar en cuenta:  $(3)^8=6561$ . Si utilizamos la simplificación con 1000 escenarios generados aleatoriamente estimamos la “probabilidad” de que cada dato censurado pertenezca a una de las poblaciones:

Dato Censurado	Prob. Población 1	Prob. Población 2	Prob. Población 3
----------------	-------------------	-------------------	-------------------

-0.4477	0.7335	0.1603	0.1062
0.4764	0.4577	0.2155	0.3268
0.8923	0.2591	0.195	0.5459
1.5217	0.0892	0.3655	0.5453
2.0699	0.015	0.5722	0.4128
1.4271	0.048	0.2654	0.6866
1.1051	0.1285	0.234	0.6375
1.2809	0.1365	0.263	0.6005

Como se puede ver en la tabla anterior, lo que hizo el método LVPLA para los 1000 escenarios generados fue mandar los datos censurados más pequeños a la población 1, los más grandes a la población 2 y los demás a la población 3.

A continuación construimos dos barreras: la primera va a separar a los datos de la población 1 de la población 3, de tal manera que todo lo que se encuentre a la izquierda se asigna a la población 1, la segunda barrera separa la población 3 de la 2, asignando todo lo que se encuentra a la derecha a la población 2, mientras lo que queda en medio de las dos barreras a la población 3.

Consideramos todos los casos que se pueden obtener, es decir, moviendo ambas barreras a través de los datos censurados, con la restricción de que la segunda barrera tiene que ser mayor o igual que la primer barrera, por lo que tenemos 45 casos, los cuales hay que maximizar.

Una vez considerados los 45 casos, nos quedamos con aquel que maximiza la función de verosimilitud, en este caso tenemos la siguiente asignación de los datos, tanto del método LVPLA, como de su simplificación:

Valor Censurado	Población Asignada por LVPLA	Población Asignada por su simplificación
-0.4477	1	1
0.4764	1	1
0.8923	3	3
1.5217	2	2
2.0699	2	2
1.4271	3	3
1.1051	3	3

1.2809	3	3
--------	---	---

Como se puede ver en ambos métodos se asignan de forma idéntica los datos censurados, por lo que los estimadores máximos verosímiles son los mismos.

### **5.5. EJEMPLO CON DOS POBLACIONES DONDE EXISTE MÁS DE UNA BARRERA**

En el ejemplo que vimos anteriormente en este capítulo de dos poblaciones, teníamos una sola barrera que separaba a los datos en dos regiones. No siempre va a ser así, podemos tener el caso, con dos poblaciones en que se separe en más de dos regiones, como el que se va a ver a continuación.

Generamos dos grupos de datos con Normal(.5,1) y Normal(0,4) cada una con 25 datos y se censuran con 50 datos generados de una Normal(.5,2), y tenemos los siguientes datos:

N(.5,1)	N(0,4)	Censura	
0.5802	3.1763	1.3634	3.4883
1.3149	-2.4499	0.0941	1.6420
-0.4060	-2.3533	2.5312	1.2544
1.0312	-1.4802	0.3020	1.6001
1.7744	-3.9509	0.2043	1.1123
2.0687	-2.0567	0.3583	-1.9490
0.2107	-2.7843	1.3438	3.3293
1.2980	-0.9918	1.6399	-0.6638
-0.8590	-1.7127	0.6931	0.0912
0.4358	-0.2281	0.3193	-2.2738
-0.2761	0.2506	0.0095	-1.3785
1.7566	-1.1174	-0.4882	-1.2661
1.7099	1.4362	1.8482	1.3600
1.3466	2.5205	-0.2980	3.4428
2.3147	0.6255	-0.2988	0.9340
1.1787	-1.9295	1.7802	-1.3603
1.3409	1.5590	1.7192	1.5386
0.4365	-0.3639	0.7388	-4.2513

-0.0215	1.7788	0.4491	2.3108
0.6800	1.7781	0.3549	1.7689
-0.8035	-0.1085	0.0307	0.2996
-0.4205	0.4553	0.3601	4.9254
-0.5957	-2.3236	0.3756	-0.0325
0.0631	2.0656	-0.1078	1.4696
1.8912	-0.4305	1.9044	-0.4167

A continuación nos quedamos con el mínimo entre los datos generados y los censurados y tenemos lo siguiente:

Población 1	Población 2	Censurados
0.5802	3.1763	0.0941
-0.4060	-2.4499	0.3020
0.2107	-2.3533	0.2043
1.2980	-1.4802	0.3583
-0.8590	-3.9509	0.3193
-0.2761	-2.0567	-0.4882
1.7099	-2.7843	-0.2980
1.1787	-0.9918	-0.2988
1.3409	-1.7127	0.3549
0.4365	2.5205	-0.1078
-0.0215	0.6255	-2.2738
-0.8035	-1.9295	-1.3785
-0.4205	1.7788	-1.2661
-0.5957	-0.1085	1.3600
1.8912	0.4553	1.5386
	-2.3236	-4.2513
	-0.4305	1.7689
		1.4696

Tenemos 15 datos que sabemos que pertenecen a la población 1, 17 a la población 2 y 18 que no sabemos a que población pertenecen. El número de posibilidades que tendría que considerar el método LVPLA son:  $2^{18}=262,144$ , por lo que utilizamos la simplificación.

Nuevamente, generamos 1000 escenarios, para la variable que nos indica si el dato censurado pertenece o no a una población y tenemos lo siguiente:

Dato Censurado	Pro. Población 1	Pro. Población 2
0.0941	0.6315	0.3685
0.3020	0.6083	0.3917
0.2043	0.5988	0.4012
0.3583	0.5922	0.4078
0.3193	0.587	0.413
-0.4882	0.6032	0.3968
-0.2980	0.5821	0.4179
-0.2988	0.5902	0.4098
0.3549	0.5668	0.4332
-0.1078	0.6345	0.3655
-2.2738	0.5335	0.4665
-1.3785	0.5512	0.4488
-1.2661	0.567	0.433
1.3600	0.4176	0.5824
1.5386	0.3688	0.6312
-4.2513	0.4585	0.5415
1.7689	0.2842	0.7158
1.4696	0.3623	0.6377

Como se puede ver los datos más grandes se asignan a la población 2, mientras que los datos más chicos se asignan a la población 1, sin embargo, el dato más chico se asigna a la población 2.

Por lo anterior se construyen dos barreras, de tal forma de que los datos que estén en medio de las dos barreras se asignan a la población 1, mientras que los datos fuera de las dos barreras, ya sea por la izquierda o la derecha se asignan a la población 2. Se deben mover ambas barreras a través de los datos censurados, siempre y cuando la segunda barrera no sea menor a la primer barrera, en este caso, tenemos 171 casos. Al hacer esto tenemos los siguientes estimadores:

$$\hat{\mu}_1 = 0.6146861, \hat{\varphi}_1 = -0.137288, \hat{\mu}_2 = -0.02602688, \hat{\varphi}_2 = 0.8258899.$$

Los estimadores de la varianza quedan como:  $\hat{\sigma}_1^2 = 0.7598940$  y  $\hat{\sigma}_2^2 = 5.216256$

A continuación calculamos las matrices observadas de Fisher, que en este caso, resultan ser:

$$I_1 = \begin{pmatrix} 0.04180794 & 0.010218511 \\ 0.010218511 & 0.03109678 \end{pmatrix} \text{ y } I_2 = \begin{pmatrix} 0.2561658 & 0.0159608 \\ 0.0159608 & 0.0307082 \end{pmatrix}.$$

Los intervalos de confianza al 95% son:

$$\mu_1 : (.213925, 1.01544), \varphi_1 : (-.48292, 1.208343), \mu_2 : (-1.0180, 96598)$$

$$\varphi_2 : (.482424, 1.1693558).$$

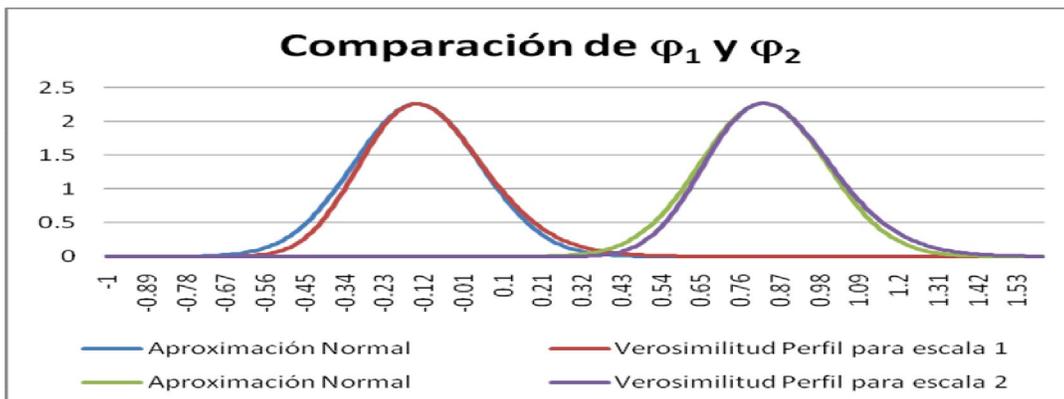
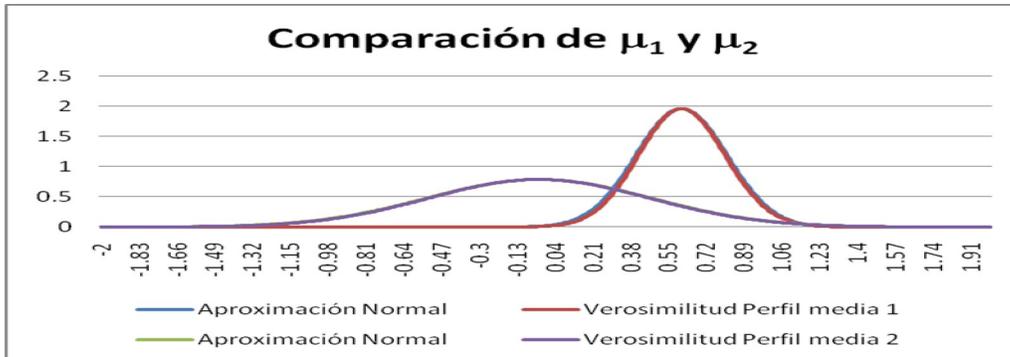
Para las varianzas tenemos:  $\sigma_1^2 : (.38066, 1.51692)$  y  $\sigma_2^2 : (2.62438, 10.3678)$ .

Todos los intervalos de confianza contienen al verdadero valor del parámetro.

Dato Censurado	Población Asignada	Población Verdadera
0.0941	1	1
0.3020	1	1
0.2043	1	1
0.3583	1	1
0.3193	1	1
-0.4882	1	1
-0.2980	1	1
-0.2988	1	1
0.3549	1	1
-0.1078	1	1
-2.2738	2	2
-1.3785	2	2
-1.2661	1	2
1.3600	2	2
1.5386	2	2
-4.2513	2	2
1.7689	2	2
1.4696	2	2

Como se puede ver todos los datos, excepto uno, fueron asignados de forma correcta, esto no se debe a la simplificación, sino como ya se había visto en el capítulo 3, que no siempre los datos se asignan de forma exacta.

Si graficamos las verosimilitudes perfil, junto con sus correspondientes verosimilitudes perfil tenemos:



Como se puede ver en las cuatro gráficas la aproximación normal sigue siendo muy buena para la verosimilitud perfil.

### **5.6. EJEMPLO CON TRES POBLACIONES DONDE EXISTEN CUATRO BARRERAS**

Tenemos tres poblaciones provenientes de distribuciones normales cada uno, con parámetros  $N(0,0.0025)$ ,  $N(0,2)$  y  $N(0,15)$ , y se censuran con una Normal  $(.5,3)$ ,

nuevamente nos quedamos con el mínimo entre estos valores. Para abreviar espacio, solo se pondrá los valores observados y los censurados:

Población 1	Población 2	Población 3	Censurados
0.0794	1.0980	3.5930	-0.1452
-0.1000	0.8555	-6.3955	-0.1455
0.0273	0.1771	-5.7833	-0.1610
-0.0542	-0.9200	-7.7182	0.0845
-0.0373	-0.8942	-3.5907	0.0078
0.0787	2.5790	4.6664	1.2800
0.0255	1.2148	-6.2912	-1.0220
0.0576	1.2579	-1.9633	-1.0352
-0.0212	-1.2563	-5.6749	-1.4452
0.0619	-1.2443	-1.2147	0.9193
0.0136	1.4552	-2.2183	1.6518
0.0194	-1.0074	4.7896	2.7753
0.0494	-1.5577	-4.9353	2.5892
0.0953	0.0697	-1.0418	-2.0410
0.0523	2.5228	2.1705	-2.1478
-0.0580	-0.1621		2.7376
-0.0261	-0.2657		-2.1237
0.0493	-0.6144		-2.3968
0.0302	1.7904		-2.0469
0.0818			2.0647
			3.6370

Tenemos 22 datos censurados, lo cual significaría que bajo el método LVPLA existen  $3^{21}=10,460,353,206$  de casos a considerar, por lo que decidimos hacer la simplificación. Tomamos 2000 escenarios y utilizamos el método LVPLA para éstos, y una vez hecho esto, tenemos la siguiente tabla con los datos censurados en forma ordenada ascendente:

Dato Censurado	Prob 1	Prob 2	Prob 3
-2.3968	0.28592708	0.06472239	0.64935053
-2.1478	0.2521664	0.08089677	0.66693682
-2.1237	0.41959231	0.48102722	0.09938047
-2.0469	0.15959362	0.64420495	0.19620142
-2.0410	0.24259678	0.6806378	0.07676542
-1.4452	0.23339858	0.62307192	0.14352949
-1.0352	0.57340418	0.34774035	0.07885547
-1.0220	0.71445026	0.21981003	0.06573971
-0.1610	0.42339605	0.29791882	0.27868513

-0.1455	0.75429309	0.07690476	0.16880215
-0.1452	0.59625493	0.15393011	0.24981496
0.0078	0.47968087	0.14833476	0.37198437
0.0845	0.09753639	0.5298154	0.37264821
0.9193	6.59E-07	0.80237855	0.19762079
1.2800	8.45E-07	0.53376426	0.46623489
1.6518	1.03E-06	0.07627876	0.92372021
2.0647	1.22E-06	0.27498027	0.72501851
2.5892	1.92E-06	0.16325821	0.83673987
2.7376	2.94E-06	0.10482853	0.89516853
2.7753	3.66E-06	0.16153005	0.83846629
3.6370	3.23E-06	0.03474909	0.96524768

Como se puede ver los datos extremos van a la población tres, los datos de la población 2 también presentan cierta separación, mientras que los datos de la población 1 están más concentrados.

A continuación se proponen cuatro barreras:

La primera barrera separa los datos pequeños de la población 3 de los datos pequeños de la población 2.

La segunda barrera separa los datos pequeños de la población 2 de los datos de la población 1.

La tercera barrera separa los datos de la población 1 de los datos grandes de la población 2.

La cuarta barrera separa los datos grandes de la población 2 de los datos grandes de la población 3.

Tenemos que mover las cuatro barreras, a través de los datos censurados, con la restricción de que la barrera 1 siempre debe estar a la izquierda de las demás barreras, la barrera 2 debe estar a la izquierda de la barrera 3 y 4 y la barrera tres a la izquierda de la barrera 4.

La asignación quedó de la siguiente forma:

Dato Censurado	Población Asignada	Población Verdadera
-2.3968	3	3
-2.1478	3	3
-2.1237	3	3
-2.0469	3	3
-2.0410	2	3
-1.4452	2	2
-1.0352	2	2
-1.0220	2	2
-0.1610	1	1
-0.1455	1	1
-0.1452	1	1
0.0078	1	1
0.0845	1	1
0.9193	2	2
1.2800	2	2
1.6518	3	2
2.0647	3	3
2.5892	3	3
2.7376	3	3
2.7753	3	3
3.6370	3	3

Sólo dos datos quedaron asignados de forma incorrecta; como ya se mencionó anteriormente esto también podría ocurrir en el método LVPLA.

Los estimadores máximo verosímiles quedan como:  $\hat{\mu}_1 = 0.020914$ ,  $\hat{\varphi}_1 = -2.95935$ ,  
 $\hat{\mu}_2 = 0.468787$ ,  $\hat{\varphi}_2 = 0.258877$ ,  $\hat{\mu}_3 = .280684$ ,  $\hat{\varphi}_3 = 1.60007$ .

Los estimadores para la varianza son:  $\hat{\sigma}_1^2 = .002689$ ,  $\hat{\sigma}_2^2 = 1.678255$  y  $\hat{\sigma}_3^2 = 24.5362$ .

Las matrices observadas de Fisher son:

$$I_1 = \begin{pmatrix} .0001258 & .0000957 \\ .0000957 & .0253535 \end{pmatrix}, I_2 = \begin{pmatrix} .097632 & .00546 \\ .00546 & .0253595 \end{pmatrix} \text{ y } I_3 = \begin{pmatrix} 1.2547 & .0615 \\ .0615 & .0349 \end{pmatrix}$$

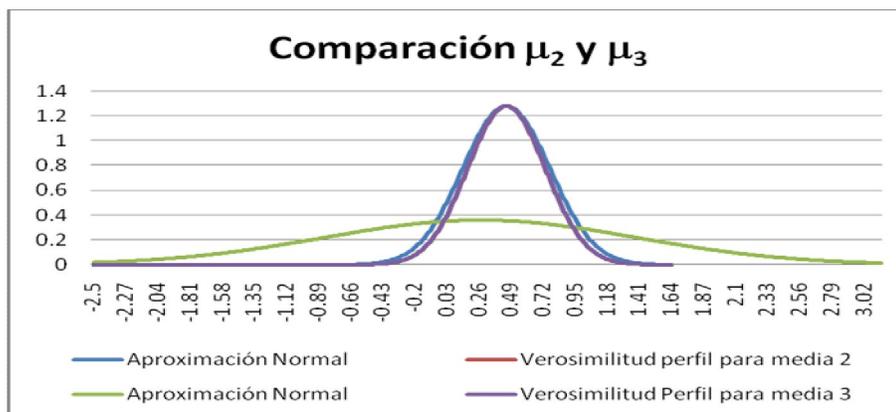
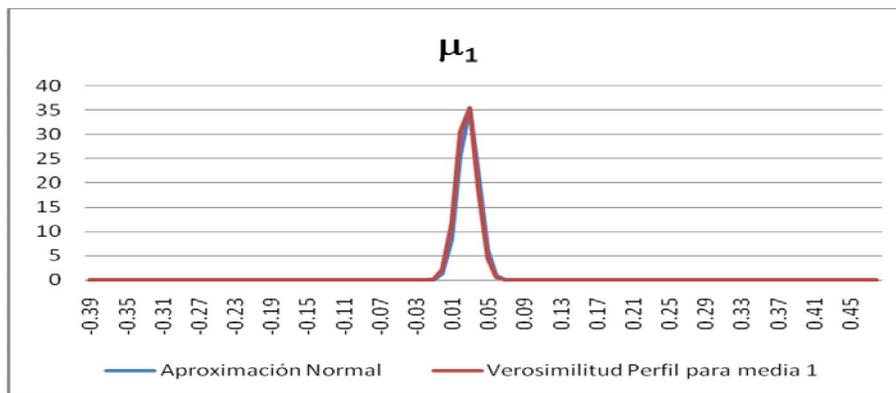
Los intervalos de confianza al 95% son:

$\mu_1 : (-.00107, .0429)$ ,  $\varphi_1 : (-3.2714, -2.6472)$ ,  $\mu_2 : (-.08337, 1.02094)$ ,  
 $\varphi_2 : (-.05346, 3.1343933)$ ,  $\mu_3 : (-1.9147, 2.4761)$  y  $\varphi_3 : (1.2329, 1.9662)$ .

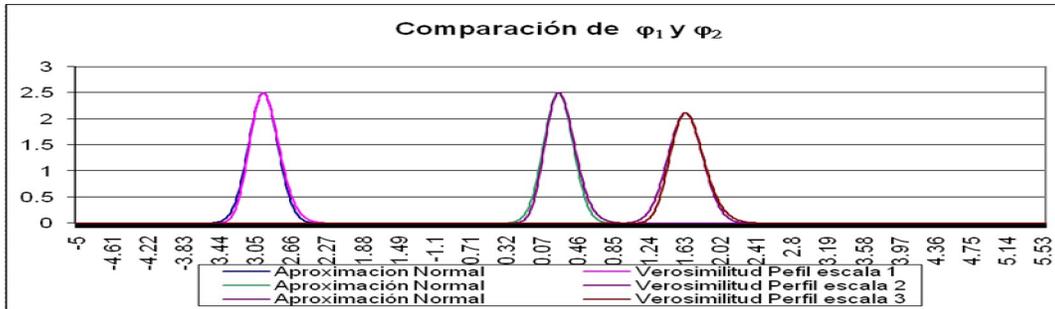
Para las varianzas tenemos:  $\sigma_1^2 : (.00144, .005019)$ ,  $\sigma_2^2 : (.89859, 3.134399)$  y  
 $\sigma_3^2 : (11.7968, 51.0327)$

Todos los intervalos de confianza contienen el verdadero valor del parámetro.

A continuación graficamos las verosimilitudes perfil con su respectiva aproximación normal, comenzando por las medias.



Las tres aproximaciones son muy buenas, además se puede apreciar que la varianza de la primera es muy pequeña, mientras que la varianza de la tercera es muy grande. La primer media no se graficó junto con las otras dos medias debido a que su distribución tiene una varianza muy chica, por lo que la moda es muy grande y no se podría apreciar bien las distribuciones para las otras medias, pero se puede ver que las tres distribuciones están alrededor del cero.



Se aprecia en la anterior gráfica, que los parámetros de escala, están muy alejados entre sí; esto es, debido a que se construyó este ejemplo, con poblaciones que tuvieran la misma media pero diferentes varianzas.

Todos los ejemplos que se construyeron en este capítulo fueron con Normales, esto es debido a la facilidad con que se generan, pero podría haberse hecho con cualquier distribución perteneciente a la familia de localización y escala.

## 5.7. ANÁLISIS DE LA ESTABILIDAD DE LA SIMPLIFICACIÓN DEL MÉTODO LVPLA

Debido a que esta simplificación del método LVPLA no utiliza todos los casos, sino solo algunos seleccionados de forma aleatoria. Puede surgir el problema de que si se escoge otra selección de casos, las probabilidades sean muy distintas a las obtenidas anteriormente y esto provoque que cambie el patrón para seleccionar los márgenes. Si esto ocurriera entonces la asignación de los datos censurados obtenida sería muy

distinta a la verdadera asignación, y por consiguiente los estimadores máximo verosímiles obtenidos, también serían distintos.

Para corroborar esto se decidió realizar el ejemplo 1 visto en la sección 5.3 donde teníamos

- 21 datos que sabemos pertenecen a la población 1.
- 17 datos que sabemos pertenecen a la población 2 y
- 12 datos censurados.

Se corrió la simplificación sólo 100 veces en el paso 1 y se obtuvo que, aunque las probabilidades obtenidas fueron distintas, en los 100 casos se obtiene que el método de simplificación asigna los valores pequeños a la población 1, mientras que los valores grandes, a la población 2.

Aunque el número de datos censurados cuya probabilidad de pertenecer a la población 1, varía mucho en las 100 corridas, como los datos más pequeños fueron asignados a la población 1, y los más grandes a la población dos, en las 100 corridas se hubiera escogido un solo margen que separara a la izquierda los datos de la población 1 de la población 2.

El caso que sería preocupante es que existiera un caso donde los datos más pequeños van a la población 2, mientras que los grandes van a la población 1, ya que la barrera que se escogería sería exactamente contraria a la que maximizaría la función de verosimilitud.

Si obtenemos un promedio de las probabilidades de cada uno de los datos pertenezca a una de las poblaciones con las 100 corridas tenemos:

Dato Censurado	Prob 1	Prob 2
-1.2465	0.58146535	0.41853465
-0.839	0.56999604	0.43000396
-0.6041	0.56472574	0.43527426
-0.3692	0.56426634	0.43573366
0.6467	0.53520693	0.46479307

0.766	0.52455941	0.47544059
0.7736	0.50835743	0.49164257
1.0898	0.4963396	0.5036604
1.2534	0.49036931	0.50963069
1.6273	0.45213267	0.54786733
2.1276	0.40856238	0.59143762
2.446	0.28747723	0.71252277

Nuevamente como se puede ver en la anterior tabla, los datos más pequeños fueron asignados a la población 1, mientras que los más grandes a la población 2, de aquí podemos seguir con el siguiente paso.

## **6. CONCLUSIONES**

De los capítulos 2 y 3 del presente documento, se deduce que el método LVPLA, simplifica no solo la estimación de los parámetros de localización y escala, sino también para obtener las matrices observadas de Fisher, que son requisito indispensable, para hacer inferencias sobre los parámetros, como obtener intervalos de confianza o realizar pruebas de hipótesis.

Lo anterior, debido a que las expresiones de las derivadas parciales quedan muy complicadas por el método PLA, y tenemos que resolver un sistema de  $3k-1$  ecuaciones con  $3k-1$  incógnitas, mientras que por el método LVPLA tenemos que resolver varios sistemas, pero éstos son de dos ecuaciones con dos incógnitas, los cuales son mucho más fáciles de manejar que el sistema anterior, sobre todo, cuando  $k$  comienza a tomar valores considerables.

Además, para el cálculo de la matriz de Fisher, que involucra una inversa, se complica numéricamente su obtención, si  $k$  es un número considerable, ya que las dimensiones de esta matriz quedan de  $3k-1$  por  $3k-1$ . Mientras que por el método LVPLA, hay que obtener  $k$  matrices observadas de Fisher de dos por dos.

El método LVPLA había sido desarrollado previamente en Contreras *et al* (2003) donde se discutió la ventaja de este método con respecto al método LVPLA. Lo que aporta este documento es una extensión del método LVPLA para el caso de datos que se encuentren bajo censura aleatoria, donde también se ve la ventaja del método LVPLA, con respecto a lo que haría el método PLA en caso de que los datos se encuentren bajo censura aleatoria.

En el capítulo 4 se aplicó el método LVPLA en un caso de literatura, utilizando distintas funciones de densidad (Valor Extremo, Weibull, Normal), y lo que se pudo apreciar es que sin importar la función de densidad que se tomará el método clasificó bajo el mismo criterio todos los datos y las funciones de supervivencia fueron distintas para ambas poblaciones en todos los casos. Bajo el método LVPLA, se pudo estimar los parámetros, además de que se encontró una forma de asignar los datos censurados a las

poblaciones. Mientras que el método PLA no asigna los datos censurados, sino que se limita a estimar la proporción de datos pertenecientes a cada población.

El único inconveniente del método LVPLA, es que, como se está suponiendo que se conoce el número de datos censurados a cada población, y en el caso de censura aleatoria, incluso que conocemos a que población pertenece cada dato censurado, es necesario “considerar” todos los casos.

El número de casos posibles también puede ser un número muy grande, sobre todo en el método LVPLA, donde el número de casos que hay que considerar es de  $k^r$ , donde “ $k$ ” es el número de poblaciones y “ $r$ ” el número de datos censurados.

Por lo anterior, en el Capítulo 5, se buscó una forma para simplificar el método LVPLA, la cual consiste en realizar el método LVLPA, sólo para un subconjunto de todos los casos posibles, para ver de que manera separa a los datos censurados el método LVPLA, con las barreras, las cuales se van moviendo para encontrar las que optimizan la verosimilitud. El número de casos que se tienen que considerar es mucho menor que si utilizáramos el método LVPLA.

El resultado anterior se obtuvo al notar que el método LVPLA con censura aleatoria clasifica los datos en cada una de las poblaciones de acuerdo con la magnitud de éstos, y en base a esto, se construyó la simplificación del método, para no tener que considerar todos los casos.

Esta simplificación también es un aporte del presente trabajo y reduce de forma significativa el número de casos a considerar y por lo tanto la estimación de los parámetros utilizando la función de verosimilitud.

Una vez resuelto este problema, el siguiente objetivo sería aplicar el método LVPLA, así como su simplificación, pero ahora, considerando variables explicativas y un esquema de una regresión.

## 7 BIBLIOGRAFÍA

Baker P, Mengersen K, and Davis G (2005) “A Bayesian Solution to reconstructing centrally censored distributions” *Journal of Agricultural Biological and Environmental Statistics*, 1, 61-84,

Campos *et al*, (sometido en *Communications*) “Inference for Mixtures of distributions of centrally censored data with partial identificación”.

Collett David (2003) *Modelling Survival Data in medical research*.

Contreras-Cristian A., Gutiérrez-Peña E., O'Reilly F. (2003): “Inference Using Latent Variables for Mixture of Distributions for Censored Data with Partial Identification” (749-759).

Crowley J. y Hu, M (1977). “Covariance Analysis of Heart transplant survival data” *Journal of American Statistical Association*, 72, 27-36

Dempster A. P., Laird, N., Rubin D. B. (1977). “Maximum Likelihood from incomplete data via de EM algorithm”. *J.R. Statist. Soc. B* 39:1-38

Diaz-Frances E, (1998) “Scientific Applications of Maximum Likelihood in Multiparametric Problems”, Ph.D. Dissertation, CIMAT, Mexico.

Greenwood M. (1926). “The natural duration of cancer. Reports on Public Health and Medical Subjects”. London: Her Majesty's Stationery Office 1926; 33:1-26.

Kaplan, E.L; Meier Paul (1958) “Nonparametric Estimation from incomplete observations”. *Journal of American Statistics Association* 53, 457-481(1958)

Lawless J. (2003) “Statistical models and methods for lifetime data”. *Wiley Series in Probability and Statistics*

Mendenhall, W., Hader, R.J, (1958) "Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data". *Biometrika* 45:504-520

Nelder, J.A., Mead, R. (1965) "A simplex method for function minimization". *Computer Journal* 7:308-313

O'Reilly F. (2003) "Significance Distributions" (8-12)

## APENDICE 1: DATOS DEL EJEMPLO DEL CAPÍTULO 4

A continuación se presentan los datos que manejamos:

Día del Transplante	Ultima vez visto	Vivo o Muerto	Rechazado	Tiempo de Vida en días
06/01/1968	21/01/1968	1	0	15
02/05/1968	05/05/1968	1	0	3
31/08/1968	17/05/1970	1	1	624
22/08/1968	07/10/1968	1	1	46
09/09/1968	14/01/1969	1	0	127
05/10/1968	08/12/1968	1	1	64
26/10/1968	07/07/1972	1	0	1350
22/11/1968	29/08/1969	1	1	280
20/11/1968	13/12/1968	1	0	23
08/02/1969	29/11/1971	1	1	1024
29/03/1969	07/05/1969	1	1	39
13/04/1969	13/04/1971	1	1	730
22/05/1969	01/04/1974	0	0	1775
16/08/1969	17/08/1969	1	0	1
03/09/1969	18/12/1971	1	1	836
14/09/1969	13/11/1969	1	1	60
16/01/1970	01/04/1974	0	0	1536
03/01/1970	01/04/1974	0	0	1549
19/05/1970	12/07/1970	1	1	54
09/05/1970	09/05/1970	1	0	0
21/05/1970	11/07/1970	1	0	51
05/01/1971	18/02/1971	1	0	44
22/02/1971	14/04/1971	1	1	51
11/08/1971	05/01/1972	0	1	147
18/08/1971	08/10/1971	1	0	51
08/11/1971	01/04/1974	1	0	875
20/11/1971	24/01/1972	1	1	65
17/03/1972	22/05/1972	1	1	66
18/05/1972	01/01/1973	1	0	228
10/06/1972	01/04/1974	0	0	660
20/08/1971	01/04/1974	0	0	955
22/09/1972	04/10/1972	1	0	12
31/05/1973	01/04/1974	0	0	305
04/02/1973	05/03/1973	1	1	29
24/02/1973	13/04/1973	1	0	48
07/03/1973	29/12/1973	1	1	297
19/05/1973	08/07/1973	1	1	50
27/04/1973	01/04/1974	0	0	339
02/03/1974	01/04/1974	0	0	30
07/08/1973	01/04/1974	0	0	237
23/09/1973	07/10/1973	1	0	14
16/10/1973	01/04/1974	0	0	167

## APENDICE 2: PROGRAMAS

(Datos en censura aleatoria en\_medio)

```
function [m1,f1,m2,f2,A,B]=censura_enmedio(X,Y,C1,C2,r)
```

```
% VARIABLES DE SALIDA
```

```
%m1 Media para la primera poblacion
```

```
%f1 Escala para la primera poblacion
```

```
%m2 Media para la segunda poblacion
```

```
%f2 Escala para la segunda poblacion
```

```
%A Matriz observada de Fisher para la primera poblacion
```

```
%B Matriz observada de Fisher para la segunda poblacion
```

```
% Variables de entrada
```

```
%X datos no censurados de la primera poblacion
```

```
%Y datos no censurados de la segunda poblacion
```

```
%C1 Extremo inferior de la region de censura
```

```
%C2 Extremo superior de la region de censura
```

```
%r Numero de Datos censurados de la mezcla
```

```
m0=mean(X); %Se obtiene un punto inicial para m1
```

```
f0=log(std(X,1)); %Se obitien un punto inicial para f1
```

```
for i=0:r %Se mueve el numero observaciones censuradas para la poblacion 1
```

```
    d=1; % Se inicializa la variable de error
```

```
    d0=[m0;f0]; % Se inicializa el vector de parametros
```

```
    while(d>.00000000001) %Mientras el error se mayor a .00000000001
```

```
        d1=d0-
```

```
inv(derivada_enmedio2(X,d0(1),d0(2),C1,C2,i))*derivada_enmedio(X,d0(1),d0(2),C1,
```

```
C2,i); % Iteraciones con el metodo de Newton
```

```
        d=norm(d1-d0); %Error
```

```
        d0=d1; % Se encuentra el siguiente punto a estimar
```

```
    end
```

```
    M1(i+1)=d0(1); %Se encuentra el optimo para cada uno de los numeros de  
observaciones censuradas
```

```
    F1(i+1)=d0(2); %Se encuentra el optimo para cada uno de los numeros de  
observaciones censuradas
```

```
    m0=M1(i+1); %Se encuentra un punto inicial para el siguiente valor
```

```
    f0=F1(i+1); %Se encuentra un punto inicial para el siguiente valor
```

```
end
```

```
m0=mean(Y); %Se obtiene un punto inicial para m1
```

```
f0=log(std(Y,1)); %Se obitien un punto inicial para f1
```

```
for i=0:r %Se mueve el numero observaciones censuradas para la poblacion 1
```

```
    d=1; % Se inicializa la variable de error
```

```

d0=[m0;f0]; % Se inicializa el vector de parametros

while(d>.00000000001) %Mientras el error se mayor a .00000000001
    d1=d0-
inv(derivada_enmedio2(Y,d0(1),d0(2),C1,C2,i))*derivada_enmedio(Y,d0(1),d0(2),C1,
C2,i); %Iteraciones con el metodo de Newton
    d=norm(d1-d0);%Error
    d0=d1; % Se encuentra el siguiente punto a estimar
end

M2(i+1)=d0(1); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas
F2(i+1)=d0(2); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas

m0=M2(i+1); %Se encuentra un punto inicial para el siguiente valor
f0=F2(i+1);%Se encuentra un punto inicial para el siguiente valor
end

for i=0:r

l(i+1)=verosimilitud_enmedio(X,M1(i+1),F1(i+1),C1,C2,i)*verosimilitud_enmedio(Y,
M2(r-i+1),F2(r-i+1),C1,C2,r-i); % Se valua la verosimilitud para cada una de las
observaciones censuradas asignadas con sus respectivos estimadores maximo
verosimiles
    L(i+1)=l(i+1)*nchoosek(r,i);
end

maximo=find(L==max(L)); % Se encuentra el maximo

m1=M1(maximo); %Estimacion de los estimadores maximo verosimiles
m2=M2(r-maximo+2); %Estimacion de los estimadores maximo verosimiles
f1=F1(maximo); %Estimacion de los estimadores maximo verosimiles
f2=F2(r-maximo+2); %Estimacion de los estimadores maximo verosimiles

A=-1*inv(derivada_enmedio2(X,m1,f1,C1,C2,maximo-1)); % Se encuentra la matriz de
fisher
B=-1*inv(derivada_enmedio2(Y,m2,f2,C1,C2,r-(maximo-1))); % Se encuentra la
matriz de fisher

function [d]=derivada_enmedio(X,m1,f1,C1,C2,r)

%VARIABLES DE SALIDA

%d Vector de derivadas

%VARIABLES DE ENTRADA

%X Datos no censurados de la poblacion

```

```

%m1 Media asignada
%f1 Escala asignada
%C1 Extremo inferior de la region de censura
%C2 Extremo inferior de la region de censura
%r Numero de Datos censurados de la mezcla

n=size(X,1);

d(1)=-(-2*sum(X)+2*n*m1)/2*exp(-2*f1)-r*exp(-f1)*(normpdf((C2-m1)*exp(-f1),0,1)-normpdf((C1-m1)*exp(-f1),0,1))/(normcdf((C2-m1)*exp(-f1),0,1)-normcdf((C1-m1)*exp(-f1),0,1));
d(2)=(X'*X-2*sum(X)*m1+n*m1*m1)*exp(-2*f1)-n*r*exp(-f1)*((C2-m1)*normpdf((C2-m1)*exp(-f1),0,1)-(C1-m1)*normpdf((C1-m1)*exp(-f1),0,1))/(normcdf((C2-m1)*exp(-f1),0,1)-normcdf((C1-m1)*exp(-f1),0,1));

d=d';

function [A]=derivada_enmedio2(X,m1,f1,C1,C2,r)

%Variables de Salida

%A Hessiano

%Variable de Entrada

%X Datos no censurados de la poblacion
%m1 Media asignada
%f1 Escala asignada
%C1 Extremo inferior de la region de censura
%C2 Extremo superior de la region de censura
%r Numero de Datos censurados de la mezcla

%Se aproxima numericamente la segunda derivada

h=.000000001;

d=derivada_enmedio(X,m1,f1,C1,C2,r);
dh=derivada_enmedio(X,m1+h,f1,C1,C2,r);
dh2=derivada_enmedio(X,m1,f1+h,C1,C2,r);

A(1,1)=(dh(1)-d(1))/h;
A(1,2)=(dh(2)-d(2))/h;
A(2,1)=A(1,2);
A(2,2)=(dh2(2)-d(2))/h;

```

(Censura No Aleatoria con tres poblaciones)

```

function [m1,f1,m2,f2,m3,f3,A,B,C]=censura_tres_pob(X,Y,Z,C,r)

%VARIABLES DE SALIDA

%m1 Media para la primera poblacion
%f1 Escala para la primera poblacion
%m2 Media para la segunda poblacion
%f2 Escala para la segunda poblacion
%m3 Media para la tercera poblacion
%f3 Escala para la tercera poblacion
%A Matriz observada de Fisher para la primera poblacion
%B Matriz observada de Fisher para la segunda poblacion
%C Matriz observada de Fisher para la tercera poblacion

%Variables de entrada

%X datos no censurados de la primera poblacion
%Y datos no censurados de la segunda poblacion
%Z datos no censurados de la tercera poblacion
%C Extremo inferior de la region de censura
%r Numero de Datos censurados de la mezcla

m0=mean(X); %Se obtiene un punto inicial para m1
f0=log(std(X,1)); %Se obtien un punto inicial para f1

for i=0:r %Se mueve el numero observaciones censuradas para la poblacion 1
    d=1; % Se inicializa la variable de error
    d0=[m0;f0]; % Se inicializa el vector de parametros
    while(d>.00000000001) %Mientras el error se mayor a .00000000001
        d1=d0-inv(derivada2(X,d0(1),d0(2),C,i))*derivada(X,d0(1),d0(2),C,i);
    %Iteracciones con el metodo de Newton
        d=norm(d1-d0); %Error
        d0=d1; % Se encuentra el siguiente punto a estimar
    end
    M1(i+1)=d0(1); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas
    F1(i+1)=d0(2); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas

    m0=M1(i+1); %Se encuentra un punto inicial para el siguiente valor
    f0=F1(i+1); %Se encuentra un punto inicial para el siguiente valor
end

m0=mean(Y); %Se obtiene un punto inicial para m2
f0=log(std(Y,1)); %Se obtien un punto inicial para f2

for i=0:r %Se mueve el numero observaciones censuradas para la poblacion 1

```

```

d=1; % Se inicializa la variable de error
d0=[m0;f0]; % Se inicializa el vector de parametros

while(d>.00000000001) %Mientras el error se mayor a .00000000001
    d1=d0-inv(derivada2(Y,d0(1),d0(2),C,i))*derivada(Y,d0(1),d0(2),C,i);
%Iteraciones con el metodo de Newton
    d=norm(d1-d0);%Error
    d0=d1; % Se encuentra el siguiente punto a estimar
end

M2(i+1)=d0(1); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas
F2(i+1)=d0(2); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas

m0=M2(i+1); %Se encuentra un punto inicial para el siguiente valor
f0=F2(i+1);%Se encuentra un punto inicial para el siguiente valor
end

m0=mean(Z); %Se obtiene un punto inicial para m3
f0=log(std(Z,1)); %Se obitien un punto inicial para f3

for i=0:r %Se mueve el numero observaciones censuradas para la poblacion 1
    d=1; % Se inicializa la variable de error
    d0=[m0;f0]; % Se inicializa el vector de parametros

    while(d>.00000000001) %Mientras el error se mayor a .00000000001
        d1=d0-inv(derivada2(Z,d0(1),d0(2),C,i))*derivada(Z,d0(1),d0(2),C,i);
%Iteraciones con el metodo de Newton
        d=norm(d1-d0);%Error
        d0=d1; % Se encuentra el siguiente punto a estimar
    end

    M3(i+1)=d0(1); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas
    F3(i+1)=d0(2); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas

    m0=M3(i+1); %Se encuentra un punto inicial para el siguiente valor
    f0=F3(i+1);%Se encuentra un punto inicial para el siguiente valor
end

l=[];
L=[];
pob1=[];
pob2=[];
pob3=[];

```

```

for i=0:r
    for j=0:r
        for k=0:r
            if(i+j+k==r) %Solo suma cuando se cumple que la suma de censuradas
asignadas a cada poblacion es el total de censuradas

L=verosimilitud(X,M1(i+1),F1(i+1),C,i)*verosimilitud(Y,M2(j+1),F2(j+1),C,j)*verosi
militud(Z,M3(k+1),F3(k+1),C,k); % Se valua la verosimilitud para cada una de
las observaciones censuradas asignadas con sus respectivos estimadores
maximo verosimiles
            L=[L;!*factorial(r)/factorial(i)/factorial(j)/factorial(k)];
            pob1=[pob1;i]; %indicadores para el numero de datos censurados
asignados a la poblacion 1
            pob2=[pob2;j]; %indicadores para el numero de datos censurados
asignados a la poblacion 2
            pob3=[pob3;k]; %indicadores para el numero de datos censurados
asignados a la poblacion 3
        end
    end
end
end

```

```

maximo1=pob1(find(L==max(L))); % Se encuentra el maximo para cada
poblacion de acuerdo con el numero de datos censurados asignados
maximo2=pob2(find(L==max(L))); % Se encuentra el maximo para cada
poblacion de acuerdo con el numero de datos censurados asignados
maximo3=pob3(find(L==max(L))); % Se encuentra el maximo para cada
poblacion de acuerdo con el numero de datos censurados asignados

```

```

m1=M1(maximo1+1); %Estimacion de los estimadores maximo verosimiles
m2=M2(maximo2+1); %Estimacion de los estimadores maximo verosimiles
m3=M3(maximo3+1); %Estimacion de los estimadores maximo verosimiles
f1=F1(maximo1+1); %Estimacion de los estimadores maximo verosimiles
f2=F2(maximo2+1); %Estimacion de los estimadores maximo verosimiles
f3=F3(maximo3+1); %Estimacion de los estimadores maximo verosimiles

```

```

A=-1*inv(derivada2(X,m1,f1,C,maximo1)); % Se encuentra la matriz de fisher
B=-1*inv(derivada2(Y,m2,f2,C,maximo2)); % Se encuentra la matriz de fisher
C=-1*inv(derivada2(Z,m3,f3,C,maximo3)); % Se encuentra la matriz de Fisher

```

```

function [d]=derivada(X,m,fi,C,R)

```

```

n=size(X,1);

```

```

Cociente=(1-normcdf((C-m)*exp(-fi)));
Cociente(find(Cociente==0))=.000001;

```

```

d(1)=-(-2*sum(X)+2*n*m)/2*exp(-2*fi)+R*(exp(-fi)*normpdf((C-m)*exp(-
fi))./Cociente);
d(2)=(X'*X-2*sum(X)*m+n*m*m)*exp(-2*fi)-n+R*((C-m).*(exp(-fi)*normpdf((C-
m)*exp(-fi)))./Cociente);
d=d';

function [A]=derivada2(X,m,fi,C,R)

h=.00000000001;

d=derivada(X,m,fi,C,R);
dh=derivada(X,m+h,fi,C,R);
dhh=derivada(X,m,fi+h,C,R);

A(1,1)=(dh(1)-d(1))/h;
A(1,2)=(dh(2)-d(2))/h;
A(2,1)=A(1,2);
A(2,2)=(dhh(2)-d(2))/h;

function [L]=verosimilitud(X,m,fi,C,R)

n=size(X,1);

Cociente=(1-normcdf((C-m)*exp(-fi)));
Cociente(find(Cociente==0))=.000001;

l=-(X'*X-2*sum(X)*m+n*m*m)/2*exp(-2*fi)-n*fi+R*(log(Cociente));
L=exp(l);

Censura Aleatoria

function [m1,f1,m2,f2,A1,A2]=censura_aleatoria(X,Y,C)

%VARIABLES DE SALIDA

%m1 Media para la primera poblacion
%f1 Escala para la primera poblacion
%m2 Media para la segunda poblacion
%f2 Escala para la segunda poblacion
%A Matriz observada de Fisher para la primera poblacion
%B Matriz observada de Fisher para la segunda poblacion

%Variables de entrada

%X datos no censurados de la primera poblacion
%Y datos no censurados de la segunda poblacion
%C Extremo inferior de la region de censura

```

```

m0=mean(X); %Se obtiene un punto inicial para m1
f0=log(std(X,1)); %Se obtien un punto inicial para f1

r=size(C,1);
p=2^r;

for j=1:r
    for i=1:p/(2)^j
        A(i,j)=0;
    end
    for i=p/(2)^j+1:p
        A(i,j)=1-A(i-p/(2)^j,j);
    end
end

%Se construye la matriz que nos indica si pertenece a o no a la poblacion

for i=1:p %Se mueve el numero observaciones censuradas para la poblacion
1
    d=1; % Se inicializa la variable de error
    d0=[m0;f0]; % Se inicializa el vector de parametros
    while(d>.00000000001) %Mientras el error se mayor a .00000000001
        d1=d0-
inv(derivada_aleatoria2(X,d0(1),d0(2),C,A(i,:)))*derivada_aleatoria(X,d0(1),d0(2)
),C,A(i,:)); %Iteraciones con el metodo de Newton
        d=norm(d1-d0); %Error
        d0=d1; % Se encuentra el siguiente punto a estimar
    end
    M1(i)=d0(1); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas
    F1(i)=d0(2); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas

    m0=M1(i); %Se encuentra un punto inicial para el siguiente valor
    f0=F1(i); %Se encuentra un punto inicial para el siguiente valor
end

m0=mean(Y); %Se obtiene un punto inicial para m2
f0=log(std(Y,1)); %Se obtien un punto inicial para f2

for i=1:p %Se mueve el numero observaciones censuradas para la poblacion
1
    d=1; % Se inicializa la variable de error
    d0=[m0;f0]; % Se inicializa el vector de parametros

    while(d>.00000000001) %Mientras el error se mayor a .00000000001

```

```

        d1=d0-
        inv(derivada_aleatoria2(Y,d0(1),d0(2),C,A(i,:)))*derivada_aleatoria(Y,d0(1),d0(2)
        ),C,A(i,:)); %Iteraciones con el metodo de Newton
        d=norm(d1-d0);%Error
        d0=d1; % Se encuentra el siguiente punto a estimar
    end

    M2(i)=d0(1); %Se encuentra el optimo para cada uno de los numeros de
    observaciones censuradas
    F2(i)=d0(2); %Se encuentra el optimo para cada uno de los numeros de
    observaciones censuradas

    m0=M2(i); %Se encuentra un punto inicial para el siguiente valor
    f0=F2(i);%Se encuentra un punto inicial para el siguiente valor
end

for i=1:p

l(i)=verosimilitud_aleatoria(X,M1(i),F1(i),C,A(i,:))*verosimilitud_aleatoria(Y,M2(p
-i+1),F2(p-i+1),C,A(p-i+1,:)); % Se valua la verosimilitud para cada una de las
observaciones censuradas asignadas con sus respectivos estimadores maximo
verosimiles
    L(i)=l(i)*nchoosek(r,sum(A(i,:)));
end

maximo=find(L==max(L)); % Se encuentra el maximo

m1=M1(maximo); %Estimacion de los estimadores maximo verosimiles
m2=M2(p+1-maximo); %Estimacion de los estimadores maximo verosimiles
f1=F1(maximo); %Estimacion de los estimadores maximo verosimiles
f2=F2(p+1-maximo); %Estimacion de los estimadores maximo verosimiles

A1=-1*inv(derivada_aleatoria2(X,m1,f1,C,A(maximo,:))); % Se encuentra la
matriz de fisher
A2=-1*inv(derivada_aleatoria2(Y,m2,f2,C,A(p+1-maximo,:))); % Se encuentra la matriz
de Fisher

function [d]=derivada_aleatoria(X,m1,f1,C,A)

%VARIABLES DE SALIDA

%d Vector de derivadas

%VARIABLES DE ENTRADA

%X Datos no censurados de la poblacion
%m1 Media asignada
%f1 Escala asignada

```

```

%C Extremo inferior de la region de censura
%A Vector que nos indica cada valor censurado si pertenece o no a la poblacion

n=size(X,1);
d(1)=-(-2*sum(X)+2*n*m1)/2*exp(-2*f1)-sum(exp(-f1)*(-normpdf((C-m1)*exp(-
f1),0,1))./(1-normcdf((C-m1)*exp(-f1),0,1)).*A');
d(2)=(X'*X-2*sum(X)*m1+n*m1*m1)*exp(-2*f1)-n*sum(exp(-f1)*(-(C-m1).*normpdf((C-
m1)*exp(-f1),0,1))./(1-normcdf((C-m1)*exp(-f1),0,1)).*A');

d=d'

function [A]=derivada_aleatoria2(X,m1,f1,C,B)

%Variables de Salida

%A Hessiano

%Variable de Entrada

%X Datos no censurados de la poblacion
%m1 Media asignada
%f1 Escala asignada
%C Extremo inferior de la region de censura
%A Vector que nos indica cada valor censurado si pertenece o no a la poblacion

%Se aproxima numericamente la segunda derivada

h=.00000000001;

d=derivada_aleatoria(X,m1,f1,C,B);
dh=derivada_aleatoria(X,m1+h,f1,C,B);
dh2=derivada_aleatoria(X,m1,f1+h,C,B);

A(1,1)=(dh(1)-d(1))/h;
A(1,2)=(dh(2)-d(2))/h;
A(2,1)=A(1,2);
A(2,2)=(dh2(2)-d(2))/h;

function [m1,f1,m2,f2,A1,A2]=censura_aleatoria_ejemplo(X,Y,C,indicador)

%VARIABLES DE SALIDA

%m1 Media para la primera poblacion
%f1 Escala para la primera poblacion
%m2 Media para la segunda poblacion
%f2 Escala para la segunda poblacion
%A Matriz observada de Fisher para la primera poblacion

```

```

%B Matriz observada de Fisher para la segunda poblacion
% Variable que nos indica la distribucion (1:Normal, 2: Valor Extremo, 3:
Weibull)

%Variables de entrada

%X datos no censurados de la primera poblacion
%Y datos no censurados de la segunda poblacion
%C Extremo inferior de la region de censura

m0=mean(X); %Se obtiene un punto inicial para m1
f0=log(std(X,1)); %Se obtien un punto inicial para f1

r=size(C,1);
p=2^r;

for j=1:r
    for i=1:p/(2)^j
        A(i,j)=0;
    end
    for i=p/(2)^j+1:p
        A(i,j)=1-A(i-p/(2)^j,j);
    end
end

%Se construye la matriz que nos indica si pertenece a o no a la poblacion

for i=1:p %Se mueve el numero observaciones censuradas para la poblacion
1
    d=1; % Se inicializa la variable de error
    d0=[m0;f0]; % Se inicializa el vector de parametros
    while(d>.00000000001) %Mientras el error se mayor a .00000000001
        d1=d0-
inv(derivada_aleatoria2_ejemplo(X,d0(1),d0(2),C,A(i,:),indicador))*derivada_ale
atoria_ejemplo(X,d0(1),d0(2),C,A(i,:),indicador); %Iteraciones con el metodo
de Newton
        d=norm(d1-d0); %Error
        d0=d1; % Se encuentra el siguiente punto a estimar
    end
    M1(i)=d0(1); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas
    F1(i)=d0(2); %Se encuentra el optimo para cada uno de los numeros de
observaciones censuradas

    m0=M1(i); %Se encuentra un punto inicial para el siguiente valor
    f0=F1(i); %Se encuentra un punto inicial para el siguiente valor
end

```

```

m0=mean(Y); %Se obtiene un punto inicial para m2
f0=log(std(Y,1)); %Se obtien un punto inicial para f2

for i=1:p %Se mueve el numero observaciones censuradas para la poblacion
1
    d=1; % Se inicializa la variable de error
    d0=[m0;f0]; % Se inicializa el vector de parametros

    while(d>.00000000001) %Mientras el error se mayor a .00000000001
        d1=d0-
        inv(derivada_aleatoria2_ejemplo(Y,d0(1),d0(2),C,A(i,:),indicador))*derivada_ale
        atoria_ejemplo(Y,d0(1),d0(2),C,A(i,:),indicador); %Iteraciones con el metodo
        de Newton
        d=norm(d1-d0);%Error
        d0=d1; % Se encuentra el siguiente punto a estimar
    end

    M2(i)=d0(1); %Se encuentra el optimo para cada uno de los numeros de
    observaciones censuradas
    F2(i)=d0(2); %Se encuentra el optimo para cada uno de los numeros de
    observaciones censuradas

    m0=M2(i); %Se encuentra un punto inicial para el siguiente valor
    f0=F2(i);%Se encuentra un punto inicial para el siguiente valor
end

for i=1:p

l(i)=verosimilitud_aleatoria_ejemplo(X,M1(i),F1(i),C,A(i,:),indicador)*verosimilitu
d_aleatoria_ejemplo(Y,M2(p-i+1),F2(p-i+1),C,A(p-i+1,:),indicador); % Se valua
la verosimilitud para cada una de las observaciones censuradas asignadas con
sus respectivos estimadores maximo verosimiles
    L(i)=l(i)*nchoosek(r,sum(A(i,:)));
end

maximo=find(L==max(L)); % Se encuentra el maximo

m1=M1(maximo); %Estimacion de los estimadores maximo verosimiles
m2=M2(p+1-maximo); %Estimacion de los estimadores maximo verosimiles
f1=F1(maximo); %Estimacion de los estimadores maximo verosimiles
f2=F2(p+1-maximo); %Estimacion de los estimadores maximo verosimiles

A1=-1*inv(derivada_aleatoria2(X,m1,f1,C,A(maximo,:))); % Se encuentra la matriz de
fisher
A2=-1*inv(derivada_aleatoria2(Y,m2,f2,C,A(p+1-maximo,:))); % Se encuentra la matriz
de Fisher

```

```

function [d]=derivada_aleatoria_ejemplo(X,m1,f1,C,A,indicador)

%VARIABLES DE SALIDA

%d Vector de derivadas

%VARIABLES DE ENTRADA

%X Datos no censurados de la poblacion
%m1 Media asignada
%f1 Escala asignada
%C Extremo inferior de la region de censura
%A Vector que nos indica cada valor censurado si pertenece o no a la poblacion

    if(indicador==1)

        n=size(X,1);
        d(1)=-(-2*sum(X)+2*n*m1)/2*exp(-2*f1)-sum(exp(-f1)*(-normpdf((C-m1)*exp(-
f1),0,1))./(1-normcdf((C-m1)*exp(-f1),0,1)).*A');
        d(2)=(X'*X-2*sum(X)*m1+n*m1*m1)*exp(-2*f1)-n*sum(exp(-f1)*(-(C-
m1).*normpdf((C-m1)*exp(-f1),0,1))./(1-normcdf((C-m1)*exp(-f1),0,1)).*A');

    elseif(indicador==2)

        n=size(X,1);
        G=sum(exp((X-m1)*exp(-f1)));
        H=sum(exp((X-m1)*exp(-f1)).*(X-m1)*exp(-f1));
        B=exp(-exp((C-m1)*exp(-f1)));
        A1=exp(-f1)*(exp((C-m1)*exp(-f1))*exp(-exp((C-m1)*exp(-f1)));
        A2=exp(-f1)*((C-m1)*exp((C-m1)*exp(-f1))*exp(-exp((C1-mu)*exp(-fi)));

        d(1)=-n*exp(-f1)+exp(-f1)*G+sum(A1/B.*A');
        d(2)=-n*exp(-f1)*(sum(X)-n*m1)+H+sum(A2/B.*A');

    elseif(indicador==3)

        n=size(X,1);
        d(1)=n/m1+sum(log(X))-exp(f1)*sum((X.^m1).*(log(X)))-
m1*sum(exp(fi)*((C.^(m1-1)).*A'));
        d(2)=n*exp(f1)*sum(X.^m1)-sum(exp(fi)*((C.^(m1)).*A'));

    end

function [A]=derivada_aleatoria2(X,m1,f1,C,B,indicador)

%Variables de Salida

%A Hessiano

```

%Variable de Entrada

%X Datos no censurados de la poblacion

%m1 Media asignada

%f1 Escala asignada

%C Extremo inferior de la region de censura

%A Vector que nos indica cada valor censurado si pertenece o no a la poblacion

%Se aproxima numericamente la segunda derivada

h=.00000000001;

d=derivada\_aleatoria(X,m1,f1,C,B,indicador);

dh=derivada\_aleatoria(X,m1+h,f1,C,B,indicador);

dh2=derivada\_aleatoria(X,m1,f1+h,C,B,indicador);

A(1,1)=(dh(1)-d(1))/h;

A(1,2)=(dh(2)-d(2))/h;

A(2,1)=A(1,2);

A(2,2)=(dh2(2)-d(2))/h;

(Se hace el método LVPLA con una muestra de los datos)

function [p]=nuevo\_metodo(X,Y,C)

tic %Se toma el tiempo

r=size(C,1); %El numero de datos censurados

A=random('Binomial',1,.5,2000,r); %Se generan los escenarios para las lambdas

m10=mean(X);

fi10=log(std(X,1));

d=ones(2,1);

for i=1:2000

    d(1)=m10;

    d(2)=fi10;

    diferencia=1;

    while(diferencia>.00000000001)

        d2=d+derivada2(X,d(1),d(2),C,A(i,:))\*derivada(X,d(1),d(2),C,A(i,:)); %Se

        maximiza para todos los escenarios

        diferencia=norm(d2-d);

        d=d2;

    end

    M1(i)=d(1);

    F1(i)=d(2);

```

end

M1=M1';
F1=F1';

m20=mean(Y);
fi20=log(std(Y,1));
d=ones(2,1);

for i=1:2000
    d(1)=m20;
    d(2)=fi20;
    diferencia=1;
    while(diferencia>.00000000001)
        d2=d+derivada2(Y,d(1),d(2),C,1-A(i,:))*derivada(Y,d(1),d(2),C,1-A(i,:));
        %Se maizmiza para todos los escenarios
        diferencia=norm(d2-d);
        d=d2;
    end
    M2(i)=d(1);
    F2(i)=d(2);
end

M2=M2';
F2=F2';

for i=1:2000

L(i)=exp(verosimilitud(X,M1(i),F1(i),C,A(i,:)))*exp(verosimilitud(Y,M2(i),F2(i),C,1-A(i,:))); %Se obtiene la verosimilitud
    L(i)=nchoosek(r,sum(A(i,:)))*L(i);
end

L=L';

G=sort(L);

L1=L(find(L>=G(1801))); %Nos quedmos con el 10 mas grande%
A1=A(find(L>=G(1801)),:); %Nos quedamos con el 10 mas grande%

for i=1:r
    p(i)=A1(:,i)*L1/sum(L1); %Se obtiene las probabilidades
end
toc

```

(Se obtienen los márgenes)

```

function [m1,f1,m2,f2,R]=margen(X,Y,C)

tic % Tomamos el tiempo

r=size(C,1)+1; %Numero de posibilidades una vez visto como separa

G=sort(C); %Se acomodan los datos

A=zeros(r,r-1);

A(1,:)=1;

for i=2:r;
    ind=find(C<=G(i-1));
    A(i,ind)=1; %Se construye las posibilidades
end

m10=mean(X); %Se saca la media
fi10=log(std(X,1)); %Se obtiene la varianza
d=ones(2,1);

for i=1:size(A,1)
    d(1)=m10;
    d(2)=fi10;
    diferencia=1;
    while(diferencia>.00000000001)
        d2=d+derivada2(X,d(1),d(2),C,A(i,:))*derivada(X,d(1),d(2),C,A(i,:)); %se
maximiza para todas las posibilidades
        diferencia=norm(d2-d);
        d=d2;
    end
    M1(i)=d(1);
    F1(i)=d(2);
end

M1=M1';
F1=F1';

m20=mean(Y);
fi20=log(std(Y,1));
d=ones(2,1);

for i=1:size(A,1)
    d(1)=m20;
    d(2)=fi20;
    diferencia=1;
    while(diferencia>.00000000001)

```

```

        d2=d+derivada2(Y,d(1),d(2),C,1-A(i,:))*derivada(Y,d(1),d(2),C,1-A(i,:)); %
Se maximiza para todas las posibilidades
        diferencia=norm(d2-d);
        d=d2;
    end
    M2(i)=d(1);
    F2(i)=d(2);
end

M2=M2';
F2=F2';

for i=1:size(A,1)

L(i)=exp(verosimilitud(X,M1(i),F1(i),C,A(i,:)))*exp(verosimilitud(Y,M2(i),F2(i),C,1
-A(i,:))); %Se obtiene la verosimilitud
    L(i)=nchoosek(r,sum(A(i,:)))*L(i);
end

L=L';

optimo=find(L==max(L)); %Se obtiene el maximo

m1=M1(optimo);
m2=M2(optimo);
f1=F1(optimo);
f2=F2(optimo);
R=A(optimo,:);

B1=derivada2(X,m1,f1,C,R) %Se obtiene la matriz de Fisher
B2=derivada2(Y,m2,f2,C,1-R) %Se obtiene la matriz de Fisher

toc %Termina el tiempo

```