



"El saber de mis hijos
hará mi grandeza"

UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

Programa de Posgrado en Matemáticas

Las Funciones de Verosimilitud Discretizada y
Restringida Perfil en la Inferencia Científica

T E S I S

Que para obtener el grado académico de:

Doctor en Ciencias
(Matemáticas)

Presenta:

Gudelia Figueroa Preciado

Directores de Tesis:

Dra. Eloísa Díaz-Francés Murguía
Dr. José Arturo Montoya Laos

Hermosillo, Sonora, México, Octubre de 2012.

Universidad de Sonora

Repositorio Institucional UNISON



**"El saber de mis hijos
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

SINODALES

Dr. Oscar Vega Amaya
Universidad de Sonora.

Dr. Román Viveros Aguilera
McMaster University, Canada.

Dr. José Arturo Montoya Laos
Universidad de Sonora.

Dra. Eloísa Díaz-Francés Murguía
Centro de Investigación en Matemáticas.

Dr. David Arthur Sprott
University of Waterloo.

Contenido

Introducción	1
1 Conceptos Básicos	7
1.1 Los razonamientos inductivo y deductivo	7
1.2 La modelación estadística de fenómenos aleatorios	8
1.2.1 El rol de los modelos estadísticos	8
1.2.2 Modelos empíricos y modelos mecanicistas	10
1.2.3 Modelos estadísticos para datos no-negativos	11
1.3 Funciones de verosimilitud importantes	11
1.3.1 La función de verosimilitud para variables aleatorias discretas	12
1.3.2 La función de verosimilitud relativa e intervalos de verosimilitud	15
1.3.3 Expansión de la log-verosimilitud relativa en una serie de Taylor	16
1.3.4 Propiedades de la función de verosimilitud.	18
1.3.5 La función de verosimilitud para variables aleatorias continuas	20
1.3.6 La función de verosimilitud para datos censurados	24
1.3.7 La función de verosimilitud para datos agrupados	25
1.3.8 La verosimilitud perfil en la estimación por separado de un parámetro de interés	26
1.4 Probabilidad de cobertura de un intervalo aleatorio	32
1.5 La importancia de reparametrizaciones adecuadas	36
1.6 La robusticidad en la estimación	37
1.7 Los cuantiles como parámetros estables	40
2 Críticas injustificadas a la función de verosimilitud	45
2.1 Verosimilitudes perfiles no-acotadas	46
2.2 La verosimilitud perfil de un parámetro umbral	48
2.3 La verosimilitud restringida: una solución a verosimilitudes no acotadas	50
2.4 Verosimilitudes perfiles de forma aplanada	51
2.5 Estimación en algunas densidades con parámetro umbral y una sin- gularidad	53
2.5.1 La distribución Weibull	53
2.5.2 La distribución lognormal	57

3	La función de verosimilitud discretizada	61
3.1	El rol de la función de verosimilitud perfil en la inferencia científica .	62
3.2	Muestra discretizada	63
3.3	La verosimilitud discretizada	66
3.3.1	Relación de la verosimilitud discretizada con otras verosimilitudes	71
3.4	La relevancia de la estadística computacional en las inferencias para muestras pequeñas	72
3.5	Distribución del estadístico de la razón de verosimilitudes para un caso no-regular	73
3.5.1	Aproximación continua a la verosimilitud para la exponencial con tiempo de vida garantizado (t.v.g.)	74
3.5.2	Verosimilitud discretizada para la exponencial con t.v.g.	75
3.5.3	Comparación entre las verosimilitudes restringida y discretizada	79
3.6	Simulaciones	81
3.6.1	Selección de los escenarios de simulación	82
3.6.2	Proceso de discretización de datos simulados	83
3.6.3	Cantidades de interés en el estudio de simulación	84
3.6.4	Resultados de las simulaciones	87
3.6.5	Resumen general sobre las simulaciones	97
4	Aplicaciones de las verosimilitudes restringida y discretizada	115
4.1	Ejemplos tipo I	117
4.1.1	Diámetros de árboles de cerezas negras	117
4.1.2	Tiempos de detección de cáncer en ratas experimentales	119
4.1.3	Índice de masa corporal en mujeres sonorenses	124
4.2	Ejemplos tipo II	128
4.2.1	Tiempos de espera en ratas para realizar una tarea de aprendizaje	128
4.3	Ejemplos tipo III	132
4.3.1	Máximos anuales del nivel del mar en Port Pirie, Australia	132
4.3.2	Datos de voltajes de rompimiento	135
4.3.3	Datos de dureza de fibras de vidrio	140
4.4	Discusión sobre ejemplos analizados	144
5	Conclusiones generales	145
A	Resultados importantes de Estadística Matemática	147
A.1	Condiciones usuales de Regularidad	147

A.2	Teorema de Máxima Verosimilitud	148
A.3	Distribución asintótica del estadístico de la razón de verosimilitudes	148
B	Algunos cálculos relevantes	149
B.1	Modelo límite para la distribución Weibull	149
B.2	La verosimilitud discretizada para el caso exponencial con t.v.g.	149
B.2.1	Distribución de $-2 \log R_D(\theta_0; \mathbf{x})$ en muestras de tamaño uno y σ conocida	150
B.2.2	Distribución de $-2 \log R_D(\mu_0; \mathbf{x})$ en muestras de tamaño n y σ conocida	151
B.2.3	Distribución de $-2 \log R_D(\mu_0; \mathbf{x})$ en muestras de tamaño n y σ desconocida	152
C	Algunos conjuntos de datos analizados	155
C.1	Diámetros de árboles de cerezas negras	155
C.2	Tiempos de detección de cáncer en ratas	155
C.3	Tiempos de espera de ratas para ejecutar tareas de aprendizaje	155
C.4	Máximos anuales del nivel del mar en Port Pirie, Australia	156
C.5	Voltajes de rompimiento de fluidos	156
C.6	Dureza de fibras de vidrio	156

Introducción

Al modelar estadísticamente fenómenos aleatorios naturales repetibles, resulta conveniente contar con procedimientos estadísticos informativos, que sean generales y a la vez sencillos. Por ello, la función de verosimilitud perfil juega un papel fundamental en esta tesis. Ésta permite resolver una gran variedad de problemas y dar respuesta a preguntas importantes que surgen al modelar un fenómeno aleatorio de interés.

Frecuentemente los modelos estadísticos considerados son multiparamétricos, por lo que se requiere estimar, por separado, aquellos parámetros que sean de interés. La función de verosimilitud perfil permite estimar, de manera sencilla, un parámetro de interés cuando se desconocen los parámetros restantes del modelo estadístico. Otros métodos que se han usado para este fin involucran a las verosimilitudes condicional y marginal, pero son muy restrictivos, porque requieren que la función de probabilidad o de densidad conjunta de la muestra tenga una estructura analítica particular. En contraste, la verosimilitud perfil es mucho más flexible, general y como se muestra en esta tesis, se comporta bien, inclusive en situaciones no regulares que usualmente se han considerado problemáticas.

En este trabajo se consideran modelos estadísticos multiparamétricos que se clasifican como no regulares, porque su densidad tiene una singularidad y un parámetro umbral, el cual representa una cota superior o inferior de la variable aleatoria en estudio. Se consideran además situaciones que involucran muestras pequeñas o moderadas. Por todo ello, no es posible aplicar los resultados asintóticos usuales de la teoría de máxima verosimilitud, pues no se cumplen los supuestos requeridos.

En este tipo de problemas pueden surgir situaciones donde no se cuente con suficiente información en los datos, para estimar un parámetro umbral. Es por ello que se muestra cómo los cuantiles son buenos parámetros a considerar, como una alternativa a la estimación del umbral. Los cuantiles son parámetros que pueden ser muy informativos sobre el fenómeno aleatorio de interés y suelen ser robustos. Esto en el sentido que generalmente son estables y no cambian mucho frente a pequeñas modificaciones en los modelos estadísticos considerados, siempre y cuando éstos sean razonables para los datos observados.

Otra característica favorable de los cuantiles es que, para muchos modelos estadísticos, juegan el rol de un parámetro de localización y por tanto son muy informativos. Al reparametrizar el modelo en términos de algún cuantil, es muy sencillo obtener su verosimilitud perfil y ésta por lo general suele ser más simétrica, lo cual

vuelve más eficiente la estimación de los parámetros del modelo. Además, cuando se desea comparar la distribución de dos poblaciones, los cuantiles son de gran utilidad. Por otra parte y como se muestra en esta tesis, la forma de la verosimilitud perfil de éstos, permite identificar situaciones donde es posible proponer un modelo estadístico más sencillo que describa adecuadamente el fenómeno de interés.

A pesar de que la verosimilitud perfil es informativa y fácil de calcular, ha sido criticada fuerte e injustificadamente en la literatura estadística. La crítica principal es que puede ser no acotada y por lo tanto no dar estimaciones razonables para los parámetros del modelo. En esta tesis se muestra que dichas críticas no tienen fundamento, ya que las fallas mencionadas son resultado de utilizar, de manera inapropiada, una aproximación continua a la función de verosimilitud, cuando no se satisfacen las condiciones necesarias para ello. Estos problemas desaparecen cuando se utiliza la verosimilitud restringida sugerida por George Barnard (1966) y propuesta después por Lawless (2003, p. 186), la verosimilitud exacta sugerida por Kalbfleisch (1985) y Lindsey (1999), al igual que la verosimilitud agrupada propuesta por Giesbrecht y Kempthorne (1976). En esta tesis se propone una variante de esta última, a la que se llamará *verosimilitud discretizada* y que también resuelve este tipo de problemas de manera muy eficiente. Esta verosimilitud se propone para casos particulares de densidades que tengan un parámetro umbral y una singularidad.

La verosimilitud discretizada puede considerarse un caso particular de la verosimilitud para datos censurados por intervalos y guarda una gran similitud con la verosimilitud agrupada y con la verosimilitud de un modelo multinomial regular. Para estas verosimilitudes los resultados asintóticos usuales se cumplen, por lo que pueden extenderse para la verosimilitud discretizada. Sin embargo, como en esta tesis el interés se centra en el comportamiento de los intervalos de estimación propuestos para los parámetros de interés, en el caso de muestras moderadas y pequeñas, se recurre a la Estadística Computacional y se estudian las coberturas de dichos intervalos a través de simulaciones que contemplan diversos escenarios de modelos no regulares, con parámetro umbral y con singularidades.

El explorar, entender y describir, bajo distintas situaciones problemáticas, el comportamiento que tiene la función de verosimilitud perfil discretizada en muestras de tamaño pequeño y moderado, permitió dar respuestas adecuadas a los problemas prácticos que se presentan en el Capítulo 4 de esta tesis. Muchos de ellos han sido presentados como situaciones problemáticas en la literatura estadística.

A continuación se detalla la forma en que se ha organizado este trabajo de tesis.

En el Capítulo 1 se describe el proceso de modelar estadísticamente un fenómeno aleatorio de interés y los papeles que juegan tanto la probabilidad como la estadística, en este proceso. Se explican dos maneras diferentes, más no excluyentes, de concebir un modelo estadístico. Se presentan los conceptos principales de la teoría de inferencia estadística que serán utilizados en los capítulos siguientes. Se describen algunos

tipos especiales de funciones de verosimilitud que se relacionan con la función de verosimilitud propuesta. Se incluyen además, algunos temas relevantes que se utilizarán, como son los métodos de eliminación de parámetros de estorbo, el uso de reparametrizaciones estabilizadoras y simetrizadoras de la verosimilitud, la importancia de los cuantiles como parámetros estables para obtener inferencias robustas, entre otros.

En el Capítulo 2 se presentan algunas críticas injustificadas que sobre la función de verosimilitud se han reportado en la literatura estadística. Los problemas mencionados generalmente se deben a utilizar una definición incorrecta de la función de verosimilitud, para el caso de variables aleatorias continuas que tienen densidades con alguna singularidad. Se muestra como la función de verosimilitud restringida sugerida por Barnard (1966), resuelve este tipo de problema de estimación, en particular cuando la densidad, aparte de una singularidad posee también un parámetro umbral. Se analiza toda la información que proporciona la forma de la función de verosimilitud perfil de un parámetro umbral y el cómo interpretarla adecuadamente. Se presentan algunos resultados asintóticos sobre algunas distribuciones de tres parámetros que poseen algún modelo límite, regular y con menos parámetros. Éstos pueden ser muy útiles para estimar por intervalos, ciertos parámetros de interés.

En el Capítulo 3 se propone y describe a la verosimilitud discretizada. Esta verosimilitud incorpora el hecho que todo instrumento de medición, en el mundo real, tiene forzosamente una precisión finita. Con ello, permite estimar eficientemente los parámetros de densidades con umbral y singularidades y por tanto resuelve muchos problemas de estimación reportados en la literatura científica. Al maximizar esta verosimilitud, generalmente los estimadores de los parámetros no suelen tener expresiones analíticas cerradas. Por ello, el comportamiento asintótico de los estimadores, para casos no regulares, no es fácil de estudiar analíticamente, ni de manera particular, ni en lo general. Sin embargo, la similitud que la verosimilitud propuesta guarda con la verosimilitud agrupada, la censurada por intervalos y la correspondiente a un modelo multinomial regular, permite deducir el comportamiento asintótico del estadístico de la razón de verosimilitudes, el cual permite asociar probabilidades de cobertura a los intervalos de verosimilitud perfil de los parámetros de interés. Estas coberturas son adecuadas en tanto el tamaño de muestra sea suficientemente grande.

Como en esta tesis, el interés principal se centra en muestras moderadas y pequeñas, se realizó un extenso estudio de simulación que permitió evaluar a partir de qué tamaño de muestra los intervalos de verosimilitud perfil de un parámetro de interés poseen coberturas razonables. Se consideraron diversos escenarios para el caso problemático, ampliamente reconocido en la literatura estadística, de la distribución Weibull de tres parámetros, la cual tiene una singularidad y un parámetro umbral. En la mayoría de los escenarios considerados se observaron buenas propiedades para la verosimilitud discretizada, inclusive en tamaños de muestras pequeños.

En este Capítulo 3 también se realiza, por vez primera, un estudio detallado sobre el comportamiento de la verosimilitud restringida, presentada por George Barnard (1966), la cual se compara con la verosimilitud discretizada propuesta en esta tesis. Se compara el desempeño de ambas verosimilitudes en escenarios de densidades con singularidades y parámetro umbral. Todas las simulaciones se efectuaron utilizando el lenguaje de programación Matlab V.7, en una computadora personal con procesador Intel(R) Core(TM) 2Duo, de 2.2GHz y con 4GB de memoria RAM.

Finalmente, en el Capítulo 4 se muestra la importancia de la verosimilitud perfil en la inferencia científica y su utilidad para dar respuesta a preguntas relevantes que surgen en los distintos problemas presentados. En ellos se observa que el utilizar la verosimilitud perfil para un parámetro de interés, así como el uso de reparametrizaciones que simetrizan la función de verosimilitud alrededor de su máximo, permite resolver de una manera integral, sencilla y clara, varios problemas que a través de la literatura se han considerado complicados. De hecho, algunos de estos ejemplos han generado controversias en cuanto a la selección del modelo más apropiado para ellos. En este capítulo se muestra cómo la forma de la verosimilitud perfil de parámetros relevantes, es crucial para la selección del modelo o modelos estadísticos más adecuados.

Las aportaciones principales de esta tesis son las siguientes:

1. Se muestra la utilidad de la función de verosimilitud perfil en la inferencia científica y la importancia de definir adecuadamente la verosimilitud en situaciones no regulares.
2. Se propone la *verosimilitud discretizada*, la cual da solución eficiente a los problemas de estimación de parámetros de interés, en el caso de densidades que tienen un parámetro umbral y singularidades. Se muestra que esta verosimilitud posee, desde muestras pequeñas, buenas coberturas para los intervalos de verosimilitud de aquellos parámetros que no delimitan el soporte de la variable aleatoria. Se encuentra también, que cuando se calculan intervalos de verosimilitud para un parámetro umbral, el tamaño de muestra necesario para obtener buenas coberturas está relacionado con el valor del parámetro de forma de la distribución no regular.
3. Se muestra que la forma de la verosimilitud perfil de un parámetro umbral es muy informativa e importante, ya que permite evaluar si un modelo más simple puede ser adecuado para describir la muestra observada. Se presenta también, la utilidad de la verosimilitud perfil de los cuantiles, ya que ésta permite analizar la robusticidad de las inferencias frente a cambios razonables en el modelo estadístico considerado.
4. Se propone la conveniencia de reparametrizar un modelo en términos de cuantiles, pues son parámetros estables que suelen simetrizar la verosimilitud, cuando éstos no corresponden a valores extremos de la distribución en estudio.

La verosimilitud perfil de éstos, permite obtener inferencias robustas, informativas y de fácil interpretación. Se muestra que bajo esta reparametrización, es muy común que los resultados asintóticos se cumplan desde muestras moderadas e inclusive pequeñas.

5. Se comparan, de una manera detallada, la verosimilitud restringida sugerida por Barnard (1966) y la verosimilitud discretizada propuesta en esta tesis. Esta comparación se efectúa a través de simulaciones para una gama muy amplia de escenarios, en los que se contrastan las coberturas de los intervalos de verosimilitud perfil de los parámetros de interés considerados. En particular se muestra que un cuantil pequeño, seleccionado adecuadamente, puede dar información similar a la que proporciona el parámetro umbral y a diferencia de éste, su intervalo de verosimilitud suele tener mejores coberturas que los intervalos asociados a un parámetro umbral.
6. Se utiliza la metodología propuesta en varios ejemplos prácticos, algunos de los cuales han sido considerados problemáticos en la literatura estadística. En éstos se comparan la verosimilitud restringida, la verosimilitud discretizada y la aproximación continua a la verosimilitud. Se discuten las semejanzas y diferencias, ventajas y desventajas que existen entre dichas funciones. Para cada uno de los ejemplos y bajo las diferentes verosimilitudes, se da respuesta a las preguntas de interés.

Capítulo 1

Conceptos Básicos

1.1 Los razonamientos inductivo y deductivo

En el mundo que nos rodea predomina la variabilidad y la incertidumbre, pero al mismo tiempo pueden observarse ciclos y patrones repetibles. La necesidad de entender nuestro entorno se manifiesta al tratar de modelar y explicar ciertos fenómenos aleatorios de interés y son las matemáticas las que proporcionan los modelos y métodos adecuados para lograrlo.

La probabilidad, una rama importante de las matemáticas, con su carácter deductivo permite cuantificar la incertidumbre en la ocurrencia de estos fenómenos aleatorios. La estadística, que ha sido llamada “*el lenguaje de la ciencia*” tiene un carácter tanto deductivo como inductivo. Ésta sirve de enlace entre la probabilidad y la naturaleza, para proponer y validar los modelos probabilísticos seleccionados, contrastándolos con datos observados del fenómeno aleatorio de interés.

La probabilidad y la estadística juegan un papel muy importante al modelar un fenómeno aleatorio de interés, aunque cada una con una participación diferente. La probabilidad proporciona los modelos matemáticos que describen el comportamiento de fenómenos aleatorios. Se considerará aquí, que un modelo de probabilidad es equivalente a una función de probabilidad o de densidad $f(x; \theta)$ de una variable aleatoria X , discreta o continua, donde el vector de parámetros θ está completamente especificado. Por eso es posible calcular la probabilidad de que la variable aleatoria se encuentre en cualquier intervalo de interés. A diferencia de un modelo de probabilidad, un modelo estadístico es una colección o familia de modelos de probabilidad $f(x; \theta)$, con θ desconocido y que tienen alguna característica en común. Algunos ejemplos de modelos estadísticos son:

1. $\mathcal{M} = \{f(x; \lambda) : f \text{ es Poisson } (\lambda)\}$,
2. $\mathcal{M} = \{f(x; \mu, \sigma, \beta) : f \text{ es Weibull } (\mu, \sigma, \beta)\}$.
3. $\mathcal{M} = \{f(x; \mu, \sigma) : f \text{ es normal con media } \mu \text{ y varianza } \sigma^2 = 1\}$

El proceso de modelar estadísticamente un fenómeno aleatorio repetible, requiere tanto de razonamiento deductivo como inductivo. Así, primeramente se plantea un modelo estadístico que se considere razonable para describir este fenómeno aleatorio de interés y posteriormente, a partir de los datos observados, se estimarán

los parámetros desconocidos. Se validarán los mejores modelos estimados, contrastándolos con los datos, para finalmente adoptar, modificar o rechazar estos modelos. En el caso de modificar o rechazar un modelo, se deberá regresar a los pasos de estimación y validación. Por tanto, se trata de un proceso iterativo tal como lo describió Box (1980), en el que cada vez se espera estar más cerca del modelo que mejor describa el fenómeno de interés. Sin embargo, debe recordarse que todo modelo matemático es una aproximación a la realidad, la cual por ser compleja nunca podrá ser descrita de manera exacta o perfecta.

Como se mencionó anteriormente, la probabilidad se rige por un razonamiento deductivo, al igual que muchas áreas de las matemáticas. Esto es, un razonamiento por medio del cual se obtiene una conclusión que se sigue de una serie de premisas o hipótesis. Esta conclusión será válida si las premisas lo son, pues la conclusión se obtiene de una cadena lógica de razonamiento. Así, los modelos de probabilidad se construyen en la mente y son una abstracción que sigue reglas lógicas y un razonamiento deductivo.

En estadística, el razonamiento que se sigue es tanto inductivo como deductivo. Es inductivo porque, dada la muestra, se sugieren varios modelos estadísticos que puedan ser apropiados, y deductivo porque las conclusiones se obtendrán a partir de las premisas del modelo elegido para los datos. La estimación de parámetros es un proceso deductivo ya que en éste se siguen reglas lógicas ya establecidas. El proceso es nuevamente inductivo si con base en la muestra observada, se valida el modelo o se sugieren cambios en el mismo. Debido a esta parte inductiva, las conclusiones obtenidas no son exactas y siempre llevan implícitas cierto grado de incertidumbre.

En resumen, puede decirse que la estadística relaciona los modelos de probabilidad con fenómenos aleatorios repetibles que surgen en la naturaleza. Ésta juega un rol muy importante en la ciencia, ya que el resolver un problema científico involucra estos dos tipos de razonamiento, que caracterizan a la Ciencia misma. Un aspecto importante es que las conclusiones obtenidas sean estables, en el sentido de que éstas no varíen mucho ante cambios razonables en el modelo o los supuestos considerados.

Antes de iniciar la siguiente sección debe aclararse que aunque en el contexto que se trabaja, sea más correcto utilizar los términos modelizar y modelización, en lugar de modelar y modelación, éstos últimos son los que se utilizarán a través de esta tesis ya que son los que, por lo general, se usan en estadística.

1.2 La modelación estadística de fenómenos aleatorios

1.2.1 El rol de los modelos estadísticos

Uno de los problemas fundamentales en la inferencia estadística consiste en encontrar un modelo adecuado que explique el proceso de generación de los datos que originan

una muestra observada; otro es poder predecir el comportamiento de observaciones futuras. El poder delimitar un conjunto de modelos para un problema particular requiere, por parte del investigador, entrenamiento y conocimiento de la disciplina, ya que deben argumentarse las razones por las cuales se incluye cierto modelo, entre los posibles candidatos a elegir, así como una justificación sobre el porqué descartar algún otro. La especificación del modelo seleccionado no es arbitraria, sino que está basada en las características del fenómeno bajo estudio y la manera en que los datos fueron obtenidos. Por ello, la selección de un modelo es, en ocasiones, más difícil que la estimación de los parámetros del modelo mismo y para elegir el más apropiado se debe contar con experiencia, que se va adquiriendo con la práctica.

Para modelar un fenómeno aleatorio de interés, se propondrá un modelo estadístico que sea lo más sencillo posible y que describa bien al mecanismo aleatorio que generó a los datos observados. Box (1980) y Sprott (2000, Section 1.3) plantean los siguientes pasos a seguir.

1. Obtener observaciones del fenómeno de interés.
2. Plantear un modelo estadístico $f(x; \theta)$ que sea razonable, dadas las características del fenómeno.
3. Verificar si es posible combinar experimentos que se hayan realizado para recabar información sobre el comportamiento de este fenómeno. Estos experimentos deberán ser homogéneos, en cuanto a que ambos estimen el mismo parámetro de interés θ .
4. Estimar el vector de parámetros θ , por medio de intervalos de estimación, así como de manera puntual.
5. Validar el modelo estadístico estimado. En caso de identificar alguna falla, modificarlo o incluso cambiarlo por otro distinto, regresando nuevamente al punto (2).

Estos pasos forman parte de un proceso iterativo que se espera converja a la elección del modelo más sencillo que describa bien a los datos observados. Un punto que consideramos importante incluir y que de acuerdo a nuestro conocimiento no ha sido mencionado de manera explícita en la literatura estadística, es el siguiente:

6. Comparar las inferencias sobre los parámetros de interés y las conclusiones obtenidas, con las que resultan de otros modelos estadísticos razonables que también se considere describen bien el fenómeno de interés.

En general se preferirán los modelos más simples que describan bien el fenómeno bajo estudio. Sin embargo, el modelo que describa bien a ciertos datos no tiene porqué ser único, ya que todo modelo es tan solo una aproximación a la realidad compleja y siempre habrá unos modelos más adecuados que otros, para explicar al fenómeno aleatorio de interés.

Un aspecto importante a considerar es el principio de parsimonia, que algunos conocen como *Navaja de Occam*, (“*Pluralitas non est ponenda sine neccesitate*”), esto es, no debe incrementarse, más allá de lo necesario, el número de entidades requeridas para explicar algo. Entre más simple sea el modelo estadístico, siempre que sea razonable y explique bien a los datos, será más fácil de entender, interpretar y usar para predecir los fenómenos de interés.

1.2.2 Modelos empíricos y modelos mecanicistas

Los modelos estadísticos pueden clasificarse como empíricos o mecanicistas. Debe aclararse que esta clasificación no es excluyente, pues un modelo propuesto puede poseer características de ambos.

Un modelo estadístico mecanicista o basado en la génesis de los datos, está descrito por una familia de funciones de probabilidad (o densidades), que se deducen del mecanismo físico que dio origen a estos datos. Estos modelos se pueden proponer cuando se posee un conocimiento claro del mecanismo que genera los datos. Los parámetros de un modelo mecanicista están fuertemente ligados al comportamiento del individuo, por ello, al estudiar procesos evolutivos se utilizan mucho este tipo de modelos.

Un ejemplo de modelo mecanicista se tiene en genética con el modelo Mendeliano de la herencia. Las leyes de Mendel proponen la existencia de ciertos factores hereditarios, hoy conocidos como genes y las reglas que los rigen. Estas reglas fueron corroboradas por Mendel a través de experimentos y explican cómo la información genética pasa de generación a generación. Por otro lado, el modelo lognormal es considerado, en muchas ocasiones, un modelo mecanicista. Este modelo se utiliza frecuentemente en biología para describir el crecimiento de un organismo, pues modela bien cambios aleatorios proporcionales a tamaños de individuos o de partículas. También se utiliza en ecología para modelar el crecimiento de poblaciones, pues éste consiste de un proceso multiplicativo.

Un modelo estadístico empírico, en cambio, es uno que describe bien el comportamiento aleatorio de los datos, se busca que tenga un buen ajuste a éstos, pero no toma en cuenta la manera en que estos datos se generaron. En ciencias aplicadas como medicina, epidemiología, climatología, etcétera, los modelos empíricos son los más utilizados, ya que el único requisito es explicar la variabilidad de los datos y no el mecanismo que los generó (Pawitan (2007, p. 5); Seber y Wild, (2003, p. 326)). Por ejemplo, se modela empíricamente cuando se supone normalidad, como es en el conjunto de datos de Darwin, donde la variable medida es la diferencia de altura entre parejas de plantas (cruzadas y auto-fertilizadas) y ésta se supone distribuida normalmente (Spratt, 2000, p. 141).

La distinción entre estos dos tipos de modelos es, en ocasiones, poco clara. Por ejemplo, un modelo que se clasifique como mecanicista puede incluir una gran canti-

dad de supuestos que pudieran considerarse empíricos, pero que simplifican el modelo y son de gran importancia para la selección del mismo. Por otra parte, un modelo que por el momento se considere empírico, con el transcurso del tiempo puede tener modificaciones y llegar a ser un modelo mecanicista.

Algo que siempre debe recordarse es que todo modelo es una simplificación de la realidad, por lo que es necesario lograr un balance entre los aspectos teóricos, físicos, biológicos, químicos, etcétera, que se incluirán en el modelo seleccionado, cuidando siempre que el modelo sea tratable desde un punto de vista matemático.

1.2.3 Modelos estadísticos para datos no-negativos

En algunas aplicaciones, la naturaleza de la variable en estudio es tal que no puede asumir o tomar valores negativos. Por ejemplo, medir el tiempo de vida de un componente, la cantidad de lluvia que cae en una región en cierto período de tiempo, la fuerza de resistencia de cierta fibra; todas ellas son variables claramente no negativas. Sin embargo, al seleccionar un modelo, si de antemano la búsqueda de éste se limita a aquellos donde el soporte de la variable, o bien sus parámetros, sean mayores o iguales que cero, pueden descartarse modelos que describan muy bien a los datos.

Un ejemplo es el modelo Gaussiano o Normal que en muchas ocasiones se utiliza para describir una variable aleatoria X no negativa, a pesar que el soporte (valores donde la densidad o función de probabilidad de la variable aleatoria es positiva) de una Normal son todos los números reales. En estos casos tan solo debe verificarse que, en el modelo seleccionado, la probabilidad que la variable tome valores negativos, $P(X < 0)$, sea despreciable. Este es un criterio razonable e importante, que será el utilizado en esta tesis.

Cuando se consideran modelos estadísticos que poseen un parámetro umbral μ , el cual representa una cota superior o inferior para la variable aleatoria, frecuentemente se sugiere restringir la búsqueda de modelos a aquellos donde $\mu \geq 0$. Esta práctica es demasiado restrictiva porque podrían descartarse buenos modelos, que a pesar de tener un umbral negativo, sean tales que la probabilidad de observar valores negativos sea muy baja. Por tanto, para los modelos estadísticos cuyo soporte sea todos los reales y se utilicen para modelar datos no negativos, por lo general se sugiere solamente pedir que esta última probabilidad sea despreciable.

1.3 Funciones de verosimilitud importantes

Una vez que se ha seleccionado un modelo estadístico para describir un fenómeno aleatorio de interés, es necesario encontrar un subconjunto pequeño de modelos probabilísticos, contenidos en esta familia, que hagan probable la muestra observada. Es decir, si la familia de modelos $f(x; \theta)$ es razonable para describir al fenómeno aleato-

rio de interés, representado por la variable aleatoria X , interesa proporcionar una región de valores de θ , dentro del espacio parametral Θ , que hagan más probables a los datos observados. De hecho se buscará también, el valor de θ que maximice la probabilidad de esta muestra.

Algunos científicos como Pierre Simon Laplace, Daniel Bernoulli y Ronald Fisher, tuvieron la idea de visualizar la función de probabilidad $f(x; \theta)$ como una función del parámetro θ y no de la variable aleatoria X . Esto sugiere considerar cuál hubiese sido el valor del parámetro que hace más probable el evento ocurrido o la muestra observada (x_1, \dots, x_n) , para esta variable aleatoria. Los valores del parámetro que hacen más creíbles los valores observados, serán preferidos o favorecidos por los científicos. En cambio, valores de los parámetros que no hagan probable los datos observados no son considerados como valores creíbles de θ . Sir Ronald A. Fisher fue el primero en formalizar matemáticamente esta idea y a este nuevo concepto lo llamó función de verosimilitud. La definió como proporcional a la probabilidad de la muestra observada y enfatizó que su naturaleza era por completo distinta a la de una probabilidad. Las definiciones para el caso de variables aleatorias tanto discretas como continuas, se presentan en las siguientes secciones.

La función de verosimilitud permite calcular la plausibilidad de cualquier valor del parámetro θ del modelo estadístico, a la luz de la muestra observada, esto es, permite ordenarlos de acuerdo a cuáles de estos valores de θ hacen más probables a los datos observados. Por ello, la verosimilitud es la cantidad relevante a considerar después de observar la muestra.

1.3.1 La función de verosimilitud para variables aleatorias discretas

El concepto de verosimilitud fue definido por primera vez en el artículo de Fisher (1921), referente al coeficiente de correlación poblacional ρ , como una cantidad proporcional a la probabilidad de que se obtenga la muestra observada. En el caso de observar la muestra $\mathbf{x} = (x_1, \dots, x_n)$, de las variables aleatorias discretas, idénticamente distribuídas, X_1, \dots, X_n , se define de la siguiente manera.

Definición 1.1 Función de verosimilitud. Sean X_1, \dots, X_n variables aleatorias discretas, idénticamente distribuídas con función de probabilidad $P_X(x; \theta)$, donde $\theta \in \Theta \subset \mathbb{R}^k$ es un vector de parámetros. Dada una realización x_1, \dots, x_n , de estas variables aleatorias, la función de verosimilitud $L : \Theta \rightarrow \mathbb{R}^+ \cup \{0\}$ se define como

$$L(\theta; x_1, \dots, x_n) = C(x_1, \dots, x_n)P(X_1 = x_1, \dots, X_n = x_n; \theta), \quad (1.3.1)$$

donde $C(x_1, \dots, x_n)$ es una función no negativa y acotada, que depende exclusivamente de los datos.

La función de verosimilitud es, por tanto, proporcional a la probabilidad conjunta de la muestra observada, pero vista como función del vector de parámetros θ . Fisher (1934), observó que la probabilidad (o densidad) conjunta de una muestra se puede

factorizar en dos partes. La primera, una función no negativa que depende exclusivamente de la muestra y la segunda, una función, también no negativa, que depende de los parámetros y de la muestra, pero a través de ciertas funciones de ella, a las que denotó como un vector de estadísticas $T(x_1, \dots, x_n)$ que son funciones de los datos observados,

$$P[X_1 = x_1, \dots, X_n = x_n; \theta] = h(x_1, \dots, x_n) g[T(x_1, \dots, x_n); \theta]. \quad (1.3.2)$$

A las estadísticas que conforman el vector T las llamó estadísticas suficientes, puesto que basta con ellas para determinar a la función de verosimilitud. Nótese que $T(x_1, \dots, x_n)$ bien podría ser la muestra misma, pero la estimación de los parámetros se simplificará notoriamente cuando la dimensión de T sea menor que n .

Obsérvese entonces que si se toma la función $C(x_1, \dots, x_n) = 1/h(x_1, \dots, x_n)$ en la Definición 1.1, entonces

$$L(\theta; x_1, \dots, x_n) = g[T(x_1, \dots, x_n); \theta].$$

La Definición 1.1 la consideró Fisher en general, sin especificar la relación de dependencia de las variables aleatorias. Sin embargo, en el caso de que estas variables sean independientes, la probabilidad conjunta será igual al producto de las probabilidades marginales correspondientes.

Ejemplo 1.1 *En cierta ciudad se registró, durante cinco años consecutivos, el número de terremotos que fue medido al menos con cierta magnitud, obteniendo los datos 3, 2, 1, 2 y 4. Si se considera un modelo estadístico Poisson para describir al número de terremotos con esta característica, entonces la función de probabilidad asociada tiene parámetro $\theta > 0$ es*

$$P[X = x; \theta] = \frac{\exp(-\theta)\theta^x}{x!}, \quad x = 0, 1, \dots$$

La función de verosimilitud para el parámetro de intensidad θ , dada la muestra observada, es

$$\begin{aligned} L(\theta; x_1, \dots, x_5) &= C(x_1, \dots, x_5) \prod_{i=1}^5 P[X_i = x_i; \theta] \\ &= (\prod_{i=1}^5 x_i!)(1/\prod_{i=1}^5 x_i!) \exp(-5\theta)\theta^{\sum x_i} \\ &= \exp(-5\theta)\theta^{\sum x_i} = g(\sum x_i; \theta). \end{aligned}$$

Nótese que $C(x_1, \dots, x_5) = \prod x_i!$ y $T(x_1, \dots, x_5) = \sum x_i$ sería el vector de estadísticas suficientes, el cual tiene solamente una componente y en este caso es unidimensional.

Como se mencionó anteriormente, la función de verosimilitud permite ordenar los valores del parámetro θ en términos de qué tan creíbles o probables hacen a la muestra observada. Dada la muestra observada (x_1, \dots, x_n) , se dice θ_1 es dos veces más

plausible que θ_2 , si el cociente $L(\theta_1; x_1, \dots, x_n)/L(\theta_2; x_1, \dots, x_n) = 2$. Para el ejemplo anterior esto significa que θ_1 hace a la muestra observada el doble de probable que si hubiese sido generada con un modelo Poisson con θ_2 . De esta manera, tiene mucho sentido el calcular un cociente de verosimilitudes. Al calcular estos cocientes se eliminarán las funciones $C(x_1, \dots, x_n)$ y $h(x_1, \dots, x_n)$, pues dependen solamente de los datos. Permanecerá, únicamente, la razón de las funciones $g[T(x_1, \dots, x_n); \theta]$ correspondientes a θ_1 y θ_2 . Por ello R. Fisher enfatizó el hecho de que la verosimilitud era proporcional a la probabilidad de la muestra observada.

Un aspecto importante es ubicar el lugar donde la función de verosimilitud alcanza su máximo. Por lo general habrá un valor de θ , a veces único, que maximiza la probabilidad de la muestra observada. Al estimador $\hat{\theta} = f(X_1, \dots, X_n)$ de θ , donde la verosimilitud alcanza su supremo se llama el *estimador de máxima verosimilitud* (emv) y es una función f de estadísticas suficientes. Se define como

$$\hat{\theta} = \arg \left\{ \sup_{\theta \in \Theta} L(\theta; x_1, \dots, x_n) \right\}.$$

Este supremo puede no ser único o también puede alcanzarse fuera del espacio parametral Θ , en cuyo caso se dirá que el emv no existe. El valor numérico que tome el emv, esto es, $\hat{\theta} = f(x_1, \dots, x_n)$ es el valor estimado por máxima verosimilitud. Nótese que el estimador es una variable aleatoria y la estimación es tan solo un punto.

En la mayoría de los casos, se tiene que la función de verosimilitud es unimodal, varias veces derivable y entonces el emv es único. En estos casos regulares, el emv se puede encontrar aplicando cálculo diferencial y para ello es más sencillo maximizar el logaritmo natural de la función de verosimilitud, a la cual se llamará la log-verosimilitud,

$$l(\theta; x_1, \dots, x_n) = \log L(\theta; x_1, \dots, x_n). \quad (1.3.3)$$

Como la función logaritmo es monótona (creciente), el supremo de $L(\theta; x_1, \dots, x_n)$ es el mismo que el de $l(\theta; x_1, \dots, x_n)$.

Ejemplo 1.1 (Cont.)

La log-verosimilitud es en este caso

$$l(\theta; x_1, \dots, x_n) = -n\theta + \sum_{i=1}^n x_i \log \theta,$$

de donde se obtiene que $\sup_{\theta \in \Theta} l(\theta; x_1, \dots, x_n)$ se alcanza en $\hat{\theta} = \bar{x}$, esto es, \bar{x} es el estimador de máxima verosimilitud de θ y $\bar{x} = 2.4$ es el estimado de máxima verosimilitud de θ . Nótese que basta conocer el valor de $T(\mathbf{x}) = \sum x_i$, pues es una estadística suficiente; no es necesario conocer los valores individuales de cada una de las observaciones.

1.3.2 La función de verosimilitud relativa e intervalos de verosimilitud

Al utilizar la función de verosimilitud completa para realizar inferencias sobre los parámetros de un modelo estadístico de interés, debe distinguirse de manera marcada de lo que es el método de máxima verosimilitud. Este último es un método que se centra en encontrar estimadores puntuales que tienen propiedades asintóticas como consistencia, insesgamiento asintótico y mínima varianza. El obtener intervalos de estimación, que reflejen la información contenida en la muestra finita observada, no tiene importancia bajo este método.

En contraste, el proceso de modelación estadística que se adopta y recomienda en esta tesis, se centra en rescatar toda la información disponible en la muestra observada, la cual estará plasmada en la función de verosimilitud completa, no solamente en su máximo. Aún cuando el tamaño de muestra sea moderado o pequeño, es posible dar intervalos de estimación para los parámetros de interés, cuantificando la incertidumbre o error que puedan tener éstos y buscando que ellos reflejen lo mejor posible la información, que sobre los parámetros, contienen los datos observados. Este es el tipo de inferencia estadística que se requiere en la Ciencia y por tanto se le llamará Inferencia Científica.

Es importante resaltar que el valor que toma la verosimilitud $L(\theta; x_1, \dots, x_n)$, para un θ en particular, no tiene un significado por sí solo, ya que solamente hace sentido al compararlo con la verosimilitud de otro valor de θ ; por ello son importantes las razones de verosimilitud. En particular, será de gran interés comparar la verosimilitud de cualquier valor de θ con respecto al emv $\hat{\theta}$, el valor que maximiza la probabilidad de la muestra observada. Esta información se encuentra en la llamada verosimilitud relativa, que es simplemente una versión estandarizada de la verosimilitud, que toma un valor de uno en el máximo y se define de la manera siguiente.

Definición 1.2 *Función de verosimilitud relativa.* Si $L(\theta; \mathbf{x})$ no es idéntica a cero, la función de verosimilitud relativa de θ , que se denota por $R(\theta; \mathbf{x})$ se define como

$$R(\theta; \mathbf{x}) = \frac{L(\theta; \mathbf{x})}{\sup_{\theta} L(\theta; \mathbf{x})} = \frac{L(\theta; \mathbf{x})}{L(\hat{\theta}; \mathbf{x})}, \quad (1.3.4)$$

donde $\hat{\theta}$ es el estimador máximo verosímil de θ .

Obsérvese que $R(\theta; \mathbf{x})$ satisface que $0 \leq R(\theta; \mathbf{x}) \leq 1$. Los valores de θ con $R(\theta; \mathbf{x})$ cercanos a uno, harán más creíble a la muestra observada, en comparación de otros con $R(\theta; \mathbf{x})$ cercano a cero.

Al graficar $R(\theta; \mathbf{x})$ como función de θ , se pueden visualizar las regiones del espacio parametral con los valores de θ que hacen más creíble la muestra observada; asimismo, se pueden identificar aquellas regiones que hacen poco probable o imposible a la muestra observada. A continuación se define un intervalo o región de verosimilitud, el cual será muy útil para estos propósitos.

Definición 1.3 Intervalo de verosimilitud. *Un intervalo o región de verosimilitud $IV(c)$ para θ , al nivel c , es el conjunto de valores del parámetro θ , tales que*

$$IV(c) = \{\theta : R(\theta; \mathbf{x}) \geq c\}, \quad (1.3.5)$$

para $c \in [0, 1]$.

Gráficamente, en el caso de que θ sea unidimensional, un intervalo o región de verosimilitud al nivel c , se obtiene al cortar horizontalmente la gráfica de $R(\theta; \mathbf{x})$ a una altura c . El intervalo de nivel $c = 1$ es la estimación de máxima verosimilitud. El intervalo o región de nivel $c = 0$ es todo el espacio parametral. Usualmente se realizan cortes a alturas como $c = 0.1465$, 0.25 y 0.03 , puesto que, en repeticiones del experimento, usualmente los intervalos resultantes suelen tener una probabilidad asociada de 0.95 , 0.90 y 0.99 , respectivamente, de incluir al verdadero valor del parámetro.

Ejemplo 1.1 (Cont.)

La función de verosimilitud relativa para el parámetro θ , será en este caso,

$$R(\theta; \mathbf{x}) = \exp(-n\theta) \exp\left(\sum x_i\right) \left(\frac{n\theta}{\sum x_i}\right)^{\sum x_i}. \quad (1.3.6)$$

Los extremos del intervalo de verosimilitud de nivel c , son las soluciones a la ecuación $R(\theta; \mathbf{x}) = c$. En la Figura 1.1 se muestran los intervalos de verosimilitud de θ , esto es, $\{\theta : R(\theta; \mathbf{x}) \geq c\}$, para valores de c de 0.03 , 0.15 y 0.25 .

En el caso de dos parámetros, donde $R(\theta; \mathbf{x})$ es una superficie en \mathbb{R}^3 , los intervalos de verosimilitud se transforman en regiones de verosimilitud. En este caso es usual graficar los contornos de verosimilitud de $R(\theta; \mathbf{x})$.

1.3.3 Expansión de la log-verosimilitud relativa en una serie de Taylor

El comportamiento de la verosimilitud alrededor del máximo verosímil es muy informativo acerca de la forma de toda la función. Si la función es plana, los datos indican que no pueden discriminar entre un rango amplio de valores de θ , y no pueden distinguir aquellos que hacen a la muestra más probable; es decir, se tiene un caso donde hay poca información en la muestra para estimar θ con buena precisión. En cambio, si la función de verosimilitud es acampanada y de poca apertura, se podrán dar intervalos cortos de estimación para el parámetro de interés y se tendrá más precisión en la estimación del parámetro desconocido.

La Estadística Matemática usa, de manera muy importante, la expansión en serie de Taylor de la log-verosimilitud (o de la log-verosimilitud relativa $r(\theta; \mathbf{x})$) para mostrar las propiedades y comportamiento asintótico de esta función. Las derivadas de la log-verosimilitud coinciden con las correspondientes de la log-verosimilitud relativa,

$$\frac{\partial^i l(\theta; \mathbf{x})}{\partial \theta^i} = \frac{\partial^i r(\theta; \mathbf{x})}{\partial \theta^i}, \quad \text{para } i \geq 1. \quad (1.3.7)$$

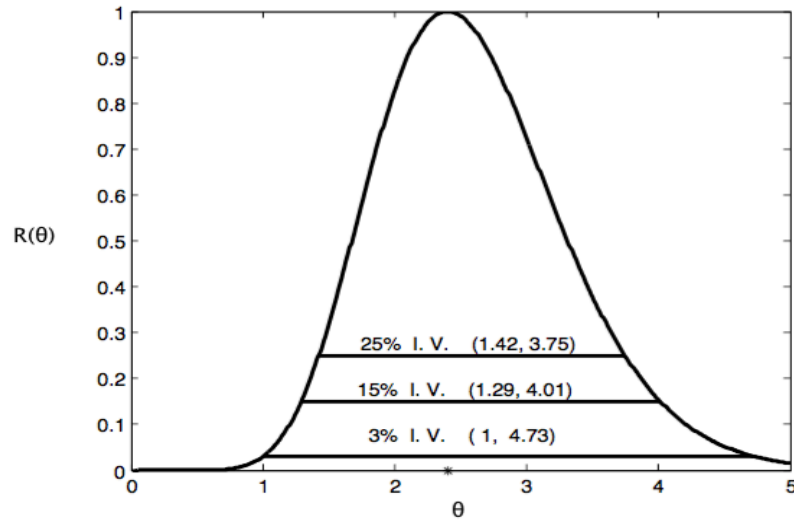


Figura 1.1: Ejemplo 1.1. Datos Poisson, verosimilitud Relativa de θ con intervalos de verosimilitud marcados para varios valores de c .

Cuando la log-verosimilitud se puede aproximar por una función cuadrática, bastará con dos cantidades para representarla, que son básicamente, la localización de su máximo y la curvatura de la función alrededor de éste. En cambio, si la log-verosimilitud no es simétrica será necesario contar con sus derivadas de mayor orden y evaluarlas en el emv $\hat{\theta}$ para describir la forma de la función completa.

Si la log-verosimilitud es diferenciable con respecto a θ , el emv se alcanza en una raíz de la ecuación que resulta de igualar a cero el vector de primeras derivadas. A este vector se le conoce como la función *score*. Esta es una función del parámetro y de los datos que se define como

$$S(\theta; \mathbf{x}) = \frac{\partial l(\theta; \mathbf{x})}{\partial \theta}. \quad (1.3.8)$$

La ecuación $S(\theta; \mathbf{x}) = 0$ se conoce como *ecuación de verosimilitud*. Esta función es de gran importancia pues permite conocer cómo cambia la verosimilitud, o la log-verosimilitud, cuando el parámetro varía. El emv encontrado como solución de la ecuación de verosimilitud debe satisfacer que

$$\left. \frac{\partial^2 l(\theta; \mathbf{x})}{\partial \theta^2} \right|_{\theta=\hat{\theta}} < 0, \quad (1.3.9)$$

para que efectivamente sea un máximo de la log-verosimilitud.

Una cantidad muy útil, denotada como *información observada* de Fisher, se basa

en evaluar el negativo de esta segunda derivada en el emv y se calcula como

$$I_{\hat{\theta}} = - \left. \frac{\partial^2 l(\theta; \mathbf{x})}{\partial \theta^2} \right|_{\theta = \hat{\theta}}. \quad (1.3.10)$$

Esta medida provee información de la curvatura de la función en $\hat{\theta}$. Así, un valor grande de $I_{\hat{\theta}}$ está asociado a una log-verosimilitud cerrada alrededor del máximo, con intervalos de verosimilitud angostos y por tanto con mayor precisión en la estimación de θ ; por eso se dice que se tiene en tal caso más información sobre θ .

Se mostrará enseguida que la expansión en serie de Taylor de $r(\theta; \mathbf{x})$ puede depender, en varias situaciones, de tan solo dos cantidades que servirían para resumir la información contenida en la función completa; éstas son el emv y la información observada de Fisher. Así, la expansión en serie de Taylor del logaritmo de la función de verosimilitud relativa alrededor de $\hat{\theta}$ se puede expresar de la manera siguiente

$$\begin{aligned} r(\theta; \mathbf{x}) &= -\frac{1}{2}(\theta - \hat{\theta})^2 I_{\hat{\theta}} + \sum_{i=3}^{\infty} \frac{1}{i!} (\theta - \hat{\theta})^i \frac{\partial^i r(\theta; \mathbf{x})}{\partial \theta^i} \\ &= -\frac{1}{2} u_{\theta}^2 + \sum_{i=3}^{\infty} \frac{(-1)^i}{i!} F_i(\hat{\theta}; \mathbf{x}) u_{\theta}^i. \end{aligned} \quad (1.3.11)$$

Obsérvese que la expansión se puede escribir como una serie de potencias en términos de la cantidad pivotal $u_{\theta} = (\theta - \hat{\theta}) \sqrt{I_{\hat{\theta}}}$ y que a partir del segundo término, los coeficientes dependen de derivadas de orden superior y están dados por

$$F_i(\hat{\theta}; \mathbf{x}) = \frac{\partial^i r(\theta; \mathbf{x})}{\partial \theta^i} I_{\hat{\theta}}^{-\frac{i}{2}}, \quad \text{para } i = 3, 4, \dots$$

En particular F_3 proporciona información acerca de la simetría de la función de verosimilitud alrededor del emv $\hat{\theta}$. La razón de expresar esta serie de Taylor en términos de u_{θ} , es porque ésta es una cantidad que juega un rol muy importante en inferencia, como se explica a detalle en la Sección 1.4.

Ejemplo 1.1 (Continuación)

Al resolver la ecuación de verosimilitud, $S(\theta; x_1, \dots, x_n) = 0$, se obtiene que el emv es $\hat{\theta} = \bar{x}$ y su estimación es 2.4; $\hat{\theta}$ es un máximo ya que $I_{\hat{\theta}} = \frac{\sum x_i}{\hat{\theta}^2} > 0$ como se observa en la Figura 1.1, donde se incluyen los intervalos del 3%, 15% y 25% de verosimilitud.

1.3.4 Propiedades de la función de verosimilitud.

Las verosimilitudes no son aditivas. La verosimilitud es una función puntual y real valuada, cuyo dominio es el espacio parametral y el contradominio son los reales no negativos. Por tanto, no hace sentido evaluar la función en la unión de dos valores de θ . Esto es, la función de verosimilitud no es una función conjuntista

como lo es la probabilidad, cuyo dominio es una sigma álgebra donde necesariamente están contenidas las uniones de eventos, y que posee además la propiedad de aditividad. A cada evento en la sigma álgebra, la probabilidad le asocia un valor en el intervalo $[0, 1]$ y la probabilidad de todo el espacio muestral \mathcal{S} es igual a uno. Por ello, tiene sentido calcular la probabilidad de la unión de dos eventos, ya que esta unión también pertenece a la sigma álgebra y se le puede asignar una probabilidad.

Por otra parte, la probabilidad es relevante antes de observar la realización de una variable aleatoria. Después de haberla observado, lo relevante será la verosimilitud, que permite hacer inferencias acerca del parámetro desconocido.

Combinación de experimentos. Cuando se tienen dos conjuntos de observaciones independientes $\mathbf{y} = (y_1, y_2, \dots, y_m)$ y $\mathbf{x} = (x_1, x_2, \dots, x_n)$ que contienen información acerca de un mismo parámetro θ , su probabilidad conjunta es tan solo el producto de sus respectivas funciones de probabilidad o de densidad; por ello, la verosimilitud basada en ambas muestras será el producto de sus verosimilitudes individuales, esto es,

$$L(\theta; \mathbf{x}, \mathbf{y}) = L(\theta; \mathbf{x})L(\theta; \mathbf{y}).$$

Lo anterior nos indica que es posible combinar experimentos donde se estime el mismo parámetro. Por otra parte, como

$$l(\theta; \mathbf{x}, \mathbf{y}) = l(\theta; \mathbf{x}) + l(\theta; \mathbf{y}),$$

al combinar dos muestras se obtendrá un estimador $\hat{\theta}$ que involucra toda la información sobre el parámetro contenida en ambas.

Invarianza funcional. Una propiedad sumamente importante de la función de verosimilitud es su invarianza respecto a reparametrizaciones uno a uno. Esta invarianza garantiza que las conclusiones del análisis estadístico serán las mismas si se reparametriza θ en términos de $\psi = g(\theta)$, siempre que g sea una transformación uno a uno de θ . Esta propiedad es muy útil pues permite escoger una reparametrización que tenga una mejor interpretación en cuanto al problema en estudio o bien seleccionar simplemente alguna que simplifique o simetrice la forma de la verosimilitud.

Por ejemplo, en el caso de la distribución exponencial con función de densidad $f(x; \theta) = \theta \exp(-\theta x)$, no importa que el parámetro de interés sea $\log \theta$, o bien la media $1/\theta$, ya que cualquiera de ellos puede considerarse. Esto es de bastante utilidad en la práctica pues en ocasiones una función de θ puede ser de mayor interés o más informativa que θ mismo.

Algo importante de mencionar es que si g es una función continua real valuada de $\Theta \subset \mathcal{R}^k \rightarrow \mathcal{R}^p$ con $p \leq k$ y $\hat{\theta}$ es el estimador de máxima verosimilitud de $\theta \in \Theta$, entonces $g(\hat{\theta})$ será el estimador de máxima verosimilitud de $g(\theta)$.

1.3.5 La función de verosimilitud para variables aleatorias continuas

El resultado de medir una cantidad física nunca es un valor único, pues siempre está implícito un error de medición. Cualquier instrumento se caracteriza por medir hasta una cierta precisión, por lo que cualquier medición es, en realidad, un intervalo continuo de valores cuya amplitud depende del error del instrumento. En instrumentos digitales el valor registrado se observa en la pantalla y el error del instrumento puede deducirse del menor dígito significativo. En los instrumentos analógicos la lectura registrada puede depender de la posición de cierto objeto, por ejemplo una aguja o divisiones con cierta escala.

Es una práctica común asumir que el error de un instrumento de medición es constante, aunque la realidad pueda ser otra. De hecho, como los instrumentos generalmente poseen un umbral a partir del cual pueden detectar una medición e igualmente un límite máximo de detección, esto implica que la precisión cambia según la magnitud de lo que se mide. Por otra parte, surgen aspectos interesantes cuando la variable aleatoria en estudio no puede ser medida directamente, pero está en función de otras variables aleatorias, que son las directamente observadas. En este caso no existe una precisión directa del instrumento de medición, sino que ésta se calcula como una función de los valores y precisiones de las variables observadas. Esto es una práctica común en experimentos de física y se conoce como propagación de errores (Shibata et al., 1995). Un ejemplo de este tipo de situación se presenta al modelar el índice de masa corporal de mujeres sonorenses, el cual se presenta en el Capítulo 4.

En esta sección se presenta la expresión correcta de la función de verosimilitud para el caso de variables aleatorias continuas. Primero se define la verosimilitud exacta presentada por Kalbfleisch (1985, p. 26), pero con la modificación que incluye Lindsey (1999) respecto a considerar que el error de medición puede ser diferente para cada una de las observaciones. Después se presenta la aproximación continua a la verosimilitud, la cual es ampliamente utilizada, aunque sólo es adecuada cuando se cumplen los supuestos que se exponen más adelante. Finalmente se presenta la verosimilitud restringida, sugerida por Barnard (1966), que se recomienda para el caso de trabajar densidades con un parámetro umbral y singularidades.

La función de verosimilitud exacta, que se define enseguida, considera que cualquier medición de una variable aleatoria continua se realiza con una precisión finita, atribuible al instrumento de medición. Por tanto, al observar $Y_i = y_i$ significa, en realidad, que Y_i puede tomar cualquier valor en el intervalo $y_i \pm h_i$, donde $2h_i$ es la precisión del instrumento de medición para la observación y_i .

Definición 1.4 *Función de verosimilitud exacta.* Sean Y_1, Y_2, \dots, Y_n variables aleatorias continuas e independientes, con función de distribución F_θ , perteneciente a una familia $\mathfrak{F} = \{F_\theta, \theta \in \Theta\}$, con $\Theta \subset \mathbb{R}^k$ y donde las funciones F_θ poseen

densidades $f(y; \theta)$. La función de verosimilitud exacta se calcula como

$$\begin{aligned} L(\theta; y_1, y_2, \dots, y_n) &= C(y_1, \dots, y_n) \prod_{i=1}^n P[y_i - h_i < Y_i \leq y_i + h_i; \theta] \\ &= F(y_i + h_i) - F(y_i - h_i), \end{aligned} \quad (1.3.12)$$

donde (y_1, y_2, \dots, y_n) denota la muestra observada, C es una función no negativa, acotada, que depende solamente de los datos y $2h_i$ es la precisión del instrumento de medición asociado con la observación y_i .

El desarrollo analítico para obtener el estimador máximo verosímil de (1.3.12) es por lo general complicado, aunque como explica Lindsey (1999) la facilidad computacional de hoy en día transforma esto en un problema menor. Esta función tiene la ventaja de que por ser un producto de probabilidades, siempre es acotada y no presenta las dificultades que pueden aparecer al utilizar la aproximación continua a la verosimilitud, que es la que mayormente se utiliza en la práctica.

Usualmente, en lugar de calcular la función de verosimilitud de la Definición 1.4 se utiliza una aproximación a ésta. Esto ocurre pues en la mayoría de las ocasiones la función de densidad es suave y regular, por lo que la contribución que la observación y_i hace a la verosimilitud, se puede calcular como

$$P[y_i - h < Y \leq y_i + h] = \int_{y_i - h}^{y_i + h} f(t; \theta) dt \stackrel{(1)}{=} 2hf(y'; \theta) \stackrel{(2)}{\approx} 2hf(y_i; \theta), \quad (1.3.13)$$

donde y' es un valor de y entre $y_i - h$ y $y_i + h$.

La igualdad (1) en la ecuación (1.3.13) se cumple cuando se satisfacen las condiciones que establece el Teorema del Valor Medio de Integrales. Ahora, si la función de densidad de Y es aproximadamente constante en el intervalo $y_i - h \leq Y \leq y_i + h$, para todo valor posible de θ (donde h no involucra a θ), se satisface la aproximación (2), de esta misma ecuación. Utilizando lo anterior se define la aproximación continua a la función de verosimilitud.

Definición 1.5 Aproximación continua a la verosimilitud. Sean Y_1, Y_2, \dots, Y_n variables aleatorias continuas, independientes, con función de distribución F_θ , que pertenece a una familia $\mathfrak{F} = \{F_\theta, \theta \in \Theta\}$, con $\Theta \subset \mathbb{R}^k$ y donde las funciones F_θ poseen densidades $f(y; \theta)$. La aproximación continua a la función de verosimilitud, la cual se denotará por L_C , se define como

$$L_C(\theta; y_1, y_2, \dots, y_n) = C(y_1, \dots, y_n) \prod_{i=1}^n f(y_i; \theta), \quad (1.3.14)$$

donde (y_1, y_2, \dots, y_n) denota la muestra observada y C es una función no negativa, acotada, que depende solamente de los datos.

A continuación se muestra, con un ejemplo sencillo, tanto el uso de la aproximación continua a la verosimilitud, como la verosimilitud exacta.

Ejemplo 1.2 *Considérese la muestra $\mathbf{y} = (y_1, y_2, \dots, y_n)$, de n observaciones de una variable aleatoria Y que se distribuye exponencialmente con tiempo medio de vida $\theta > 0$. La función de densidad está dada por*

$$f(y; \theta) = \frac{1}{\theta} \exp(-y/\theta) I_{[0, \infty)}(y), \quad (1.3.15)$$

y la función de distribución correspondiente es

$$F(y; \theta) = 1 - \exp(-y/\theta) I_{[0, \infty)}(y). \quad (1.3.16)$$

Supóngase que las observaciones y_1, \dots, y_n , fueron registradas con un instrumento de medición con precisión $2h$. La verosimilitud exacta será

$$L(\theta; \mathbf{y}) = \prod_{i=1}^n \left\{ \exp \left[- \left(\frac{y_i - h}{\theta} \right) \right] - \exp \left[- \left(\frac{y_i + h}{\theta} \right) \right] \right\} \quad (1.3.17)$$

$$= \exp \left(- \sum_{i=1}^n \frac{y_i}{\theta} \right) \left[\exp \left(\frac{h}{\theta} \right) - \exp \left(- \frac{h}{\theta} \right) \right]^n. \quad (1.3.18)$$

Igualando a cero la función Score y encontrando la raíz de esta ecuación, se obtiene que el emv de θ es

$$\hat{\theta} = 2h \left[\log \left(\frac{\bar{y} + h}{\bar{y} - h} \right) \right]^{-1}. \quad (1.3.19)$$

Por otra parte, cuando se utiliza la aproximación continua a la verosimilitud,

$$L_C(\theta; \mathbf{y}) = \frac{1}{\theta^n} \exp \left(- \frac{1}{\theta} \sum_{i=1}^n y_i \right), \quad (1.3.20)$$

se obtiene que el estimador de máxima verosimilitud de θ obtenido con la aproximación continua es $\hat{\theta}_a$, dado por

$$\hat{\theta}_a = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}.$$

Supóngase ahora que se cuenta con una sola observación de la muestra y_1, \dots, y_n , de esta distribución exponencial. La aportación que esta observación hace a la verosimilitud será igual a la probabilidad

$$P[y - h < Y \leq y + h] = F(y + h; \theta) - F(y - h; \theta).$$

A continuación se muestra la razón de esta probabilidad y de la densidad evaluada en esa única observación, misma que servirá para comparar la aproximación continua a la verosimilitud con la verosimilitud exacta,

$$\frac{\exp \left[- \left(\frac{y-h}{\theta} \right) \right] - \exp \left[- \left(\frac{y+h}{\theta} \right) \right]}{\left(\frac{1}{\theta} \right) \exp(-y/\theta)(2h)} = \frac{\exp(k) - \exp(-k)}{2k} = 1 + \frac{k^2}{3!} + \frac{k^4}{5!} + \dots$$

donde $k = h/\theta$ es el cociente de h , la mitad de la precisión del instrumento de medición, y θ , el parámetro desconocido. La aproximación será buena cuando k sea pequeño, esto es, cuando el instrumento tenga buena precisión para la magnitud de lo que se desee estimar. En este ejemplo basta que $k \leq 0.1$ para que en una gráfica, las curvas de la aproximación continua a la verosimilitud y la verosimilitud exacta sean indistinguibles. Nótese que en (1.3.19), cuando $h \rightarrow 0$ entonces el emv de θ converge a \bar{y} .

Es ocasiones, cuando se utiliza la aproximación continua a la verosimilitud, puede darse que para ciertas combinaciones de modelos y conjuntos de datos, exista una trayectoria en el espacio de parámetros, en la cual la verosimilitud tiende a infinito. Esto no necesariamente se debe a que la probabilidad de los datos sea grande en esa región del espacio parametral, sino que resulta de utilizar de manera inadecuada la aproximación continua a la verosimilitud, como explican Meeker y Escobar (1998). En estos casos, usar la aproximación continua puede conducir a dificultades numéricas y puede enmascarar un máximo local asociado al valor de la estimación máximo verosímil.

Muchas de las críticas y supuestas fallas del enfoque de verosimilitud, radican en el hecho de que se olvida que la Definición 1.5 es tan solo una aproximación a la verosimilitud. En particular, debe tenerse cuidado de utilizarla cuando la función de densidad posea una singularidad, como menciona Lindsey (1999). Tal es el caso de las densidades Weibull y lognormal de tres parámetros; la verosimilitud hereda esta singularidad y en ciertas regiones puede tomar un valor infinito. Edwards (1992, p. 167) por su parte comenta que el concepto de distribución continua es en realidad una ficción y que en ocasiones puede causar ciertos problemas. Una alternativa es utilizar lo que aquí se llamará la *verosimilitud restringida*. Ésta fue sugerida por Barnard (1966) y descrita en Lawless (2003, p. 186) y en Montoya et al. (2009) y se define a continuación.

Definición 1.6 *Función de verosimilitud restringida* Sean Y_1, Y_2, \dots, Y_n variables aleatorias continuas, independientes, con función de distribución F_θ , que pertenece a una familia $\mathfrak{F} = \{F_\theta, \theta \in \Theta\}$, con $\Theta \subset \mathbb{R}^k$. La densidad asociada $f(y; \theta)$ depende de los parámetros $\theta = (\mu, \lambda)$, donde μ es un parámetro umbral que representa una cota inferior de Y , λ es el vector de parámetros restantes y $f(y; \theta)$ tiene asociada una singularidad en $y = \mu$. La función de verosimilitud restringida, que se denotará por L_R , se calcula como

$$L_R(\mu, \lambda; \mathbf{y}) = C(y_1, \dots, y_n) \prod_{i=1}^n f(y_i; \mu, \lambda) I_{(-\infty, y_{(1)}-h]}(\mu), \quad (1.3.21)$$

donde (y_1, y_2, \dots, y_n) denota a la muestra observada, C es una función no negativa, acotada, que depende solamente de los datos y $2h$ es la precisión del instrumento de medición asociado con la observación más pequeña $y_{(1)}$.

En el Capítulo 2 se analizará en detalle el uso de la verosimilitud restringida, la cual se utiliza en todas las simulaciones y ejemplos presentados en esta tesis.

En las siguientes secciones se presentan dos funciones de verosimilitud importantes que surgen en diversos tipos de estudios, cuando en éstos existen ciertas restricciones que no permiten obtener una medición ‘exacta’. Una es la verosimilitud para muestras censuradas y la otra es la verosimilitud para datos agrupados. Ambas, de cierta manera, están relacionadas con la verosimilitud discretizada que se presenta en el Capítulo 3, razón por la que se explica detalladamente su cálculo.

1.3.6 La función de verosimilitud para datos censurados

En áreas como la Medicina y la Industria, es común observar a n pacientes hasta poder detectar una cierta enfermedad o bien monitorear n componentes durante cierto tiempo, hasta registrar una falla. Este tipo de experimentos da lugar a mediciones particularmente interesantes en cuanto a lo que es posible o no observar y se detallan a continuación.

En el caso de experimentos donde se mide un tiempo de falla T en n componentes, es común que éstos sólo se observen durante un intervalo de tiempo $[0, t_0]$, el cual generalmente se determina antes de comenzar el estudio. Si k componentes fallan antes de t_0 , se registra el tiempo de falla t_i , asociado a éstos. Cuando la función de densidad es continua, diferenciable y no oscila mucho, la aportación que estas observaciones hacen a la verosimilitud será igual a evaluarlas en la densidad del modelo estadístico considerado, $f(t_i; \theta)$. Así, si en un tiempo t_0 hay $n - k$ componentes que todavía no han fallado y sólo se sabe que su tiempo de falla ocurrirá forzosamente en el intervalo (t_0, ∞) , a estas observaciones se les llamará *censuradas por la derecha*, con un límite común t_0 . Esto ocurre seguido en ensayos clínicos, donde puede haber individuos que permanecen aún vivos en el tiempo t_0 en que el estudio finaliza, o bien, si en ese momento se encuentran libres de la enfermedad que se estaba registrando. La aportación, que estas observaciones censuradas, hacen a la verosimilitud será precisamente la probabilidad de que la variable aleatoria T se encuentre en el intervalo (t_0, ∞) , que es igual a $[1 - F_T(t_0; \theta)]$.

También puede ocurrir que el instrumento de medición tenga un límite inferior de tiempo de detección t_0 , de manera que cuando registra el valor t_0 , solamente se sabe que la falla ocurrió en el intervalo $[0, t_0]$ y no se puede distinguir de su valor exacto. En este caso la contribución a la verosimilitud, es nuevamente la probabilidad asociada de caer en este intervalo y es $F_T(t_0; \theta) - F_T(0; \theta)$. A este tipo de medición se le conoce como *censurada por la izquierda*.

Finalmente, un tipo de censura muy importante, conocida como *censura por intervalo*, resulta cuando solamente se conoce el intervalo (a, b) en el que ocurre la falla. En ese caso, la aportación que la observación hace a la verosimilitud es $F_T(b; \theta) - F_T(a; \theta)$. Un caso particular de esta situación es, precisamente, cuando se registra una observación utilizando un instrumento de medición con precisión finita $2h$, pues al registrar la observación t_i , solamente se sabe que la variable T tomó un valor en el intervalo $(t_i - h, t_i + h]$. Así, la aportación de esta observación a

la verosimilitud será $F_T(t_i + h; \theta) - F_T(t_i - h; \theta)$. Entonces, la verosimilitud exacta contempla el caso de observaciones censuradas por intervalo, todos ellos de longitud $2h$.

La función de verosimilitud considerará las aportaciones particulares de todas las observaciones, sean o no censuradas. Por ejemplo, si como se mencionó, se tienen k datos para los cuales se observaron tiempos de falla t_i , $i = 1, \dots, k$, y hay $n - k$ observaciones restantes de las que sólo se sabe fallarán después de un tiempo t_0 , en el que el estudio finalizó, la verosimilitud se calculará de la manera siguiente,

$$L(\theta; t_1, \dots, t_k, n) = C(t_1, \dots, t_k, t_0) \prod_{i=1}^k f(t_i; \theta) [1 - F(t_0; \theta)]^{n-k}.$$

Por otra parte, para aquellos casos donde todas las observaciones fueron censuradas por intervalo y se sabe que sólo se registró el valor medio t_i de los intervalos $(t_i - h, t_i + h]$, la verosimilitud resultante coincidirá con la ecuación (1.3.12) de la verosimilitud exacta presentada en la Definición 1.4.

1.3.7 La función de verosimilitud para datos agrupados

La verosimilitud para datos agrupados ocurre frecuentemente cuando las observaciones se clasifican como pertenecientes a algún intervalo, de entre un número k de ellos, los cuales pueden ser de la forma $I_j = [a_{j-1}, a_j)$, $j = 1, \dots, k$ y donde $0 = a_0 < a_1 < \dots < a_k = \infty$. Si las observaciones provienen de una variable aleatoria continua Y no negativa con función de distribución $F(y; \theta)$ y estas observaciones se agrupan en intervalos, de manera que la muestra observada de ocurrencias es (d_1, \dots, d_k) , donde $n = \sum_{i=1}^k d_i$, entonces este evento puede modelarse con una distribución multinomial para un vector aleatorio $Z = (Z_1, \dots, Z_k)$ con función de probabilidad

$$P(Z_1 = d_1, Z_2 = d_2, \dots, Z_k = d_k) = \frac{n!}{d_1! \dots d_k!} \pi_1(\theta)^{d_1} \dots \pi_k(\theta)^{d_k},$$

donde las probabilidades dependen de la distribución original de la variable continua Y , la cual depende de θ ,

$$\pi_j(\theta) = P[a_{j-1} \leq Y < a_j] = F(a_j; \theta) - F(a_{j-1}; \theta).$$

Entonces la función de verosimilitud de θ puede escribirse con base en este modelo multinomial,

$$L(\theta; d_1, \dots, d_k) = C(d_1, \dots, d_k) \prod_{j=1}^k [F(a_j; \theta) - F(a_{j-1}; \theta)]^{d_j}, \quad (1.3.22)$$

donde C es una función no negativa, acotada y que no depende de θ .

La log verosimilitud es

$$l(\theta; d_1, \dots, d_k) = \ln C(d_1, \dots, d_k) + \sum_{j=1}^k d_j \ln \pi_j(\theta).$$

Si las $\pi_j(\theta)$, $j = 1, \dots, k + 1$ son diferenciables con respecto a θ , entonces el estimador máximo verosímil de θ puede encontrarse como solución de

$$\sum_{j=1}^k \frac{d_j}{\pi_j(\theta)} \frac{\partial}{\partial(\theta)} \pi_j(\theta) = 0, \quad (1.3.23)$$

sujeto a la restricción de que $\sum_{j=1}^k \pi_j(\theta) = 1$.

Nótese que la fórmula de la verosimilitud presentada (1.3.22) es similar a la mostrada en (1.3.12), que corresponde a la de la verosimilitud exacta, sólo que en esta última la amplitud de sus intervalos está ligada a la precisión del instrumento de medición.

Cuando se utilizan verosimilitudes agrupadas, el cálculo de los estimadores máximo verosímiles no es sencillo, por lo que generalmente se hace uso de métodos numéricos para maximizar la función $l(\theta; d_1, \dots, d_k)$.

1.3.8 La verosimilitud perfil en la estimación por separado de un parámetro de interés

En los modelos estadísticos que tienen un sólo parámetro, el enfoque de verosimilitud, antes descrito, permite realizar inferencias sobre este parámetro, de una manera sencilla. Sin embargo, bajo cualquier enfoque, la situación se complica si el número de parámetros aumenta. Considérese un modelo estadístico donde θ , es un vector k dimensional de parámetros que puede expresarse en la forma $\theta = (\lambda, \psi)$, donde λ es el parámetro unidimensional de interés, y ψ es un vector de parámetros restantes, a los que suele llamarse incidentales o de estorbo, con dimensión $(k - 1)$. Todos los parámetros en θ pueden ser de gran importancia pero, para una situación particular, λ es el parámetro sobre el cual se desea hacer inferencias.

Kalbfleisch y Sprott (1970), describen varios métodos que se han usado con el fin de estimar un parámetro de interés, en presencia de parámetros de estorbo. En algunos de estos casos, para realizar inferencias en estos modelos multiparamétricos, la estructura analítica de la función de probabilidad o de densidad conjunta de la muestra, debe poder factorizarse en dos términos: uno que aisle al parámetro de interés y otro que involucre los parámetros restantes. Esta factorización se puede lograr condicionando en estadísticas suficientes o auxiliares, la probabilidad conjunta de la muestra. Así, la verosimilitud del parámetro de interés será proporcional al factor que aisle este parámetro y éste puede ser una probabilidad marginal o condicional, razón por la que heredará su nombre.

En el pasado, las verosimilitudes condicionales o marginales han sido preferidas para estimar por separado parámetros de interés, en presencia de otros de estorbo, debido a que son proporcionales a probabilidades genuinas, que se obtienen a partir de la probabilidad conjunta de la muestra observada (véase Sprott, 2000, Secciones 4.2 y 4.3). Para algunos modelos resultan tener además expresiones sencillas, lo cual facilita las inferencias sobre el parámetro de interés. Sin embargo, resulta válido usarlas solamente en aquellos casos donde la información que guarde el factor restante sobre el parámetro de interés sea despreciable.

Así pues, para utilizar la verosimilitud condicional cuando se tiene un vector de parámetros $\theta = (\lambda, \psi)$, donde λ es el parámetro de interés, se necesita contar con un estadístico $T(y_1, \dots, y_n)$ que sea suficiente minimal para el parámetro de estorbo ψ , para cualquier valor fijo del parámetro de interés λ . Es necesario que la función de probabilidad o de densidad de la muestra pueda factorizarse como,

$$f(\mathbf{y}; \lambda, \psi) = f_1(t; \lambda, \psi) f_2(\mathbf{y}; \lambda|t),$$

donde la probabilidad condicional $f_2(\mathbf{y}; \lambda|t)$ no depende de ψ porque $T(\mathbf{y})$ es suficiente para este parámetro. Así, la verosimilitud de θ será proporcional a estos dos factores y por tanto también se podrá factorizar su verosimilitud correspondiente, en dos verosimilitudes que serán proporcionales a cada uno de estos factores,

$$L(\lambda, \psi; \mathbf{y}) = L_{res}(\lambda, \psi; t) L_{cond}(\lambda; \mathbf{y}),$$

donde

$$L_{res}(\lambda, \psi; t) \propto f_1(t; \lambda, \psi) \quad \text{y} \quad L_{cond}(\lambda; \mathbf{y}) \propto f_2(\mathbf{y}; \lambda|t). \quad (1.3.24)$$

La verosimilitud condicional de λ , se denota aquí como $L_{cond}(\lambda; \mathbf{y})$ y es la que se utiliza para realizar inferencias acerca de λ , cuando no se tiene conocimiento del parámetro ψ . Está basada en la distribución condicional de la muestra $\mathbf{y} = (y_1, \dots, y_n)$ dado $T(y)$ y la utilidad de estas inferencias dependerá de hasta qué punto el factor $L_{res}(\lambda, \psi; t)$, la función de verosimilitud condicional residual correspondiente, no contenga mucha información sobre λ (Sprott, 2000, p. 50). No siempre es sencillo evaluar qué tanta información guarda, sobre el parámetro de interés, esta verosimilitud residual. Tampoco es posible encontrar, para todos los modelos estadísticos y parámetros de interés, una estadística que sea suficiente para los parámetros de estorbo restantes.

Por otra parte, para poder utilizar la verosimilitud marginal, la función de densidad o de probabilidad conjunta de la muestra debe ser tal que se pueda factorizar de la siguiente manera

$$f(\mathbf{y}; \lambda, \psi) = f_1(t; \lambda) f_2(\mathbf{y}; \lambda, \psi|t), \quad (1.3.25)$$

donde ahora $T(y)$ es una estadística auxiliar para el parámetro de estorbo ψ , esto es, una estadística tal que su distribución está por completo especificada y no depende

del parámetro ψ , sino solamente del parámetro de interés λ . En este caso el factor que se usará para realizar inferencias sobre el parámetro de interés será la densidad marginal de la estadística auxiliar.

De esta manera, la verosimilitud global se puede factorizar en dos términos,

$$L(\lambda, \psi, \mathbf{y}) = L_m(\lambda; t) L_{res}(\lambda, \psi; \mathbf{y}),$$

donde

$$L_m(\lambda; t) \propto f_1(t; \lambda) \text{ y } L_{res}(\lambda, \psi; \mathbf{y}) \propto f_2(\mathbf{y}; \lambda, \psi|t). \quad (1.3.26)$$

$L_m(\lambda; t)$ denota aquí la verosimilitud marginal de λ , y es proporcional a la densidad marginal de la estadística auxiliar $T(y)$. Esta verosimilitud es la que se utiliza para estimar por separado el parámetro de interés λ . La función $L_{res}(\lambda, \psi; \mathbf{y})$ representa la verosimilitud marginal residual. Las inferencias hechas con la verosimilitud marginal serán buenas en tanto $L_{res}(\lambda, \psi; \mathbf{y})$ no contenga mucha información sobre el parámetro de interés.

Un procedimiento más general, sumamente informativo y práctico, para estimar por separado un parámetro de interés en presencia de otros de estorbo, es la verosimilitud perfil, la cual no requiere que la probabilidad o densidad conjunta de la muestra posea una estructura analítica especial, aspecto que suele ser muy restrictivo. La verosimilitud perfil se obtiene de una manera muy sencilla, ya que para cada valor fijo del parámetro de interés, se maximiza la verosimilitud con respecto a los $k - 1$ parámetros de estorbo restantes. Esto se repite para un conjunto de valores del parámetro de interés, en un intervalo dado. En ocasiones se puede obtener una expresión analítica cerrada para la verosimilitud perfil del parámetro de interés, pero cuando esto no es posible, se maximiza la verosimilitud global utilizando procedimientos numéricos de optimización. La verosimilitud perfil es el procedimiento central que se utilizará en esta tesis para realizar inferencias acerca de los parámetros de interés en un modelo estadístico. La dimensión del parámetro de interés suele ser menor o igual a tres, aunque en esta tesis se considerará solamente de una dimensión.

El primer uso que conocemos de este tipo de verosimilitud lo dan Hood y Koopmans (1953), quienes le llaman verosimilitud concentrada y la utilizaron para reducir la dimensión de la función de verosimilitud global de todos los parámetros y facilitar así su maximización. Por otra parte, Box y Cox (1964) utilizaron la verosimilitud perfil del parámetro de forma en sus modelos de transformaciones. Sin embargo, es hasta 1970, en el artículo de de Kalbfleisch y Sprott, donde se le presenta como un método general para estimar un parámetro de interés en presencia de parámetros de estorbo y se le da el nombre de *verosimilitud maximizada o perfil*.

Consideremos nuevamente el vector de parámetros $\theta = (\lambda, \psi)$, donde el parámetro de interés es λ . Para calcular la verosimilitud perfil de λ , se fijan valores de este parámetro en una rejilla fina, usualmente estos valores son equidistantes, y para cada valor fijo $\lambda = \lambda_0$, se maximiza la log-verosimilitud con respecto de los parámetros

restantes. Con ello se obtienen valores de los parámetros de estorbo que maximizan la verosimilitud, restringidos al valor fijo λ_0 . A estos valores $\hat{\psi}(\lambda_0; \mathbf{y})$ se les llamará los estimadores restringidos de máxima verosimilitud de ψ . Este procedimiento se repite para todos los valores de λ en la rejilla elegida. Así, la verosimilitud perfil de λ puede definirse como

Definición 1.7 *La función de verosimilitud perfil.* Sea $L(\theta; \mathbf{y})$ la función de verosimilitud para $\theta = (\lambda, \psi)$. La verosimilitud perfil de λ , que se denotará por $L_p(\lambda; \mathbf{y})$, se define como

$$L_p(\lambda; \mathbf{y}) = \sup_{\psi|\lambda} L(\lambda, \psi; \mathbf{y}) = L(\lambda, \hat{\psi}(\lambda; \mathbf{y}); \mathbf{y}), \quad (1.3.27)$$

donde $\hat{\psi}(\lambda; \mathbf{y})$ es el estimador de máxima verosimilitud restringido de ψ , que se obtiene maximizando la función de verosimilitud para un valor fijo de λ .

Esto es, $L_p(\lambda; \mathbf{y})$ se obtiene sustituyendo el parámetro de estorbo ψ por $\hat{\psi}(\lambda; \mathbf{y})$ en la función de verosimilitud global $L(\lambda, \psi; \mathbf{y})$, como se verá en detalle en el Ejemplo 1.3. El estimador de máxima verosimilitud de λ , esto es, el primer componente de la pareja $(\hat{\lambda}, \hat{\psi})$ que maximiza la verosimilitud global $L(\lambda, \psi)$ es el mismo que maximiza la verosimilitud perfil $L_p(\lambda; \mathbf{y})$.

La razón del nombre de verosimilitud perfil puede comprenderse considerando un escenario con pocas dimensiones. Para ello supóngase que λ y ψ son dos parámetros unidimensionales, entonces la función de verosimilitud global será una superficie en \mathbb{R}^3 . Observando desde un punto lejano y perpendicular al eje correspondiente al parámetro de interés, el contorno que se observe de esta superficie, proyectado en un plano, corresponderá a la verosimilitud perfil, como se muestra más adelante, en la Figura 1.2(a).

La verosimilitud perfil relativa $R_p(\lambda; \mathbf{y})$ es la versión estandarizada de $L_p(\lambda; \mathbf{y})$ y se calcula como

$$R_p(\lambda; \mathbf{y}) = \frac{L_p(\lambda; \mathbf{y})}{L(\hat{\lambda}, \hat{\psi}; \mathbf{y})}. \quad (1.3.28)$$

A continuación se muestra cómo calcular la verosimilitud perfil en un modelo Weibull de dos parámetros.

Ejemplo 1.3 *Sea $\mathbf{y} = (y_1, \dots, y_n)$ la muestra observada de n variables aleatorias independientes e idénticamente distribuidas como Weibull de mínimos con parámetros (σ, β) , donde σ es un parámetro de escala y β es un parámetro de forma. Su función de densidad está dada por*

$$f(y; \beta, \sigma) = \frac{\beta}{\sigma} \left(\frac{y}{\sigma}\right)^{\beta-1} \exp\left[-\left(\frac{y}{\sigma}\right)^\beta\right] I_{[0, \infty)}(y), \quad \sigma > 0, \quad \beta > 0. \quad (1.3.29)$$

La función de verosimilitud correspondiente es

$$L(\sigma, \beta; \mathbf{y}) \propto \prod_{i=1}^n \frac{\beta}{\sigma} \left(\frac{y_i}{\sigma}\right)^{\beta-1} \exp\left[-\left(\frac{y_i}{\sigma}\right)^\beta\right] I_{(0,\infty)}(\beta) I_{(0,\infty)}(\sigma) \quad (1.3.30)$$

$$= \frac{\beta^n}{\sigma^{n\beta}} \exp\left(-\frac{1}{\sigma^\beta} \sum_{i=1}^n y_i^\beta\right) \prod_{i=1}^n y_i^{\beta-1} I_{(0,\infty)}(\beta) I_{(0,\infty)}(\sigma). \quad (1.3.31)$$

El logaritmo de la verosimilitud es

$$l(\sigma, \beta; \mathbf{y}) = n \log \beta - n\beta \log \sigma + (\beta - 1) \sum_{i=1}^n \log y_i - \frac{1}{\sigma^\beta} \sum_{i=1}^n y_i^\beta.$$

Las ecuaciones de verosimilitud se obtienen igualando las derivadas parciales a cero. De las dos ecuaciones, analíticamente solamente puede resolverse

$$\frac{\partial l(\sigma, \beta; \mathbf{y})}{\partial \sigma} = -n + \frac{\sum_{i=1}^n y_i^\beta}{\sigma^\beta} = 0,$$

de donde se encuentra explícitamente el estimador de máxima verosimilitud restringido de σ , que está dado por

$$\hat{\sigma}(\beta, \mathbf{y}) = \left(\frac{\sum_{i=1}^n y_i^\beta}{n}\right)^{1/\beta}. \quad (1.3.32)$$

Con ello, la verosimilitud perfil de β tiene la siguiente expresión analítica

$$\begin{aligned} L_p(\beta; \mathbf{y}) &= L_p(\beta, \hat{\sigma}(\beta, \mathbf{y}); \mathbf{y}) \\ &= \frac{\beta^n}{[\hat{\sigma}(\beta, \mathbf{y})]^{n\beta}} \exp\left\{-\frac{1}{[\hat{\sigma}(\beta, \mathbf{y})]^\beta} \sum_{i=1}^n y_i^\beta\right\} \prod_{i=1}^n y_i^{\beta-1} I_{(0,\infty)}(\beta) I_{(0,\infty)}(\sigma). \end{aligned}$$

En la superficie mostrada en la Figura 1.2(a) puede observarse la verosimilitud relativa para los parámetros de escala y forma para un modelo Weibull de dos parámetros. En esta figura también se incluye, proyectada en el plano $\sigma = 16$, la verosimilitud perfil relativa del parámetro β , la cual también puede observarse en la Figura 1.2(b).

La Figura 1.3(a) muestra algunos contornos de la superficie de verosimilitud relativa para los parámetros (σ, β) . En ésta, la curva punteada representa la trayectoria asociada a la verosimilitud perfil de β , proyectada sobre estos contornos; los puntos de esta trayectoria tienen coordenadas $[\hat{\sigma}(\beta, \mathbf{y}), \beta]$ para valores de β en el intervalo $[3, 6]$. En contraste, para la verosimilitud perfil de σ no se puede proporcionar una expresión analítica cerrada, ya que tampoco es posible obtenerla para el emv restringido $\hat{\beta}(\sigma, \mathbf{y})$. Por ello, la verosimilitud perfil de σ sólo es posible calcularla de manera numérica, asignando valores a σ en una rejilla regular, digamos en el intervalo $[13, 16.2]$, donde para cada valor fijo de σ , se maximiza, sobre β , la verosimilitud $L(\beta, \sigma; \mathbf{y})$. Esta verosimilitud perfil de σ puede observarse en la Figura 1.3(b).

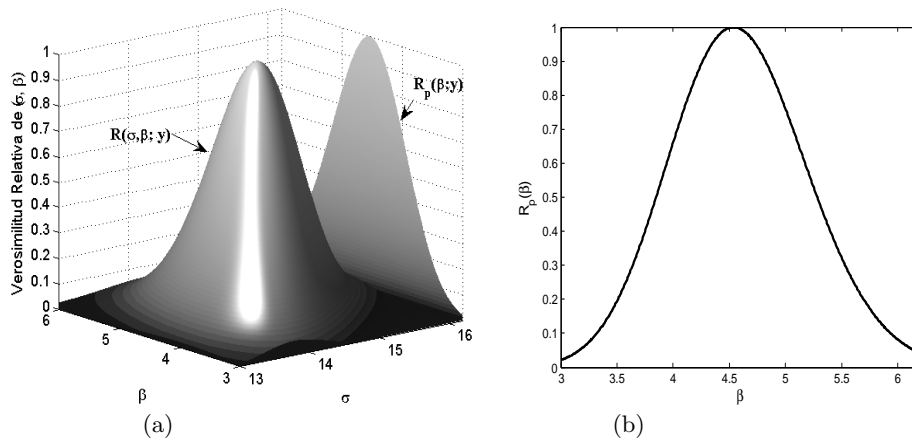


Figura 1.2: Ejemplo 1.4: (a) Superficie de verosimilitud relativa para σ y β , (b) Verosimilitud perfil relativa de β .

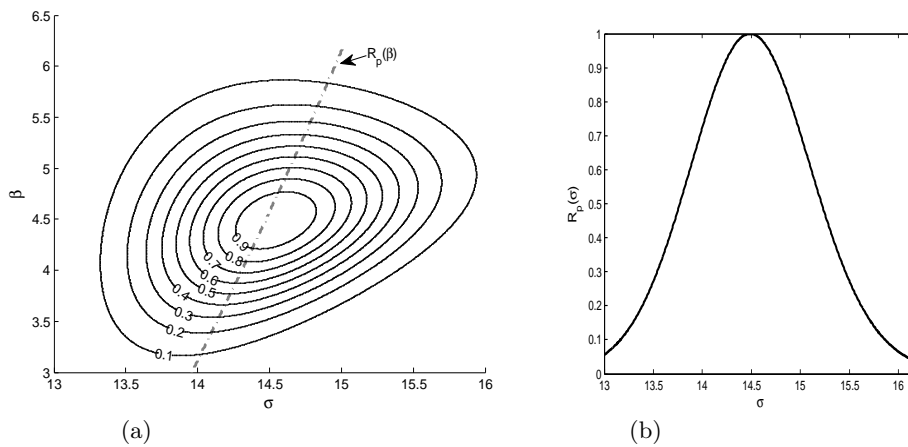


Figura 1.3: Ejemplo 1.4: (a) Contornos de verosimilitud de σ y β , (b) Verosimilitud perfil relativa de σ .

Se han discutido pues, tres maneras de estimar un parámetro de interés en presencia de parámetros de estorbo. Existen situaciones en la que las inferencias obtenidas serán similares, independientemente de la verosimilitud utilizada, como muestra Montoya (2008, pp. 32-34) en el ejemplo referente a los datos de ovejas citados por Fisher (1941), donde se obtienen inferencias similares con la verosimilitud perfil y la verosimilitud condicional. Esta última resulta sencilla de utilizar debido a la estructura, que en ese caso, tiene la función de verosimilitud. Así pues, habrá ocasiones en que será posible aislar al parámetro de interés, sin perder información relevante en la verosimilitud condicional residual, como muestra Sprott (2000, pp. 49-50), en el ejemplo donde estudia la diferencia entre dos distribuciones Poisson. Sin embargo, en otras situaciones esto no será posible de conseguir, como se muestra para el conjunto de datos ECMO (véase Fig. 4.2 Sprott, 2000, p. 55), donde la información referente al parámetro de interés no puede separarse, adecuadamente, del parámetro

de estorbo.

La ventaja de las verosimilitudes marginales y condicionales sobre la verosimilitud perfil, es que cuando resulta posible calcularlas, corresponden a una probabilidad genuina de observar los datos. Sin embargo, no siempre es obvio cómo factorizar la probabilidad conjunta o cómo encontrar estadísticas suficientes o auxiliares para los parámetros de estorbo (Pawitan, 2007, p. 279). Por otra parte, debe siempre recordarse que las verosimilitudes condicional y marginal solamente toman en cuenta un factor de la probabilidad conjunta de la muestra y por tanto, debe tenerse mucho cuidado de no perder información relevante del parámetro de interés, que se encuentre en el otro factor.

La verosimilitud perfil, en cambio, es intuitiva pues para cada valor fijo del parámetro de interés proporciona una cota superior para la verosimilitud que puede obtenerse, con los parámetros de estorbo. Es más flexible en cuanto a la eliminación de parámetros de estorbo, pues no depende de que el modelo a analizar posea una estructura especial; por ello, las inferencias obtenidas con la verosimilitud perfil son más generales. Es también muy informativa, incluso en el caso de contar con muestras pequeñas; por ello el papel tan importante que juega en la inferencia científica. Por ejemplo, en situaciones donde el parámetro de interés es un parámetro umbral, la forma de la verosimilitud perfil es crucial para las inferencias e interpretaciones que se realicen, ya que su forma puede inclusive sugerir un modelo más sencillo para describir los datos.

Otro aspecto importante de la verosimilitud perfil, es la facilidad de realizar inferencias acerca de algún otro parámetro de interés, aprovechando la propiedad de invarianza funcional de la verosimilitud. Por otra parte, como se verá en la siguiente sección, la verosimilitud perfil permite también, de una manera muy sencilla, evaluar la robusticidad en la estimación. Ahora, aunque al utilizarla pueden presentarse situaciones donde no sea posible encontrar expresiones analíticas para los estimadores de los parámetros en estudio, es generalmente sencillo implementar procesos de optimización numérica para obtener la verosimilitud perfil del parámetro o parámetros de interés.

Debe aclararse, sin embargo, que en cualquiera de las verosimilitudes mencionadas, el cálculo se complicará si el número de parámetros a estimar es muy grande. En el caso de la verosimilitud perfil, sólo cuando la dimensión del parámetro de interés sea menor o igual que cinco, las regiones de verosimilitud serán aproximadamente regiones de confianza (Barndorff-Nielsen y Cox, 2001).

1.4 Probabilidad de cobertura de un intervalo aleatorio

En la Sección 1.3.2 se definió el concepto de intervalo de verosimilitud y se explicó que en este intervalo siempre se encuentra el emv, aunque no necesariamente está en

el centro del mismo. Esto es, el intervalo no es necesariamente simétrico alrededor del estimador del máxima verosimilitud. Los intervalos de verosimilitud son muy informativos y si se cuenta con varios de ellos, es posible reproducir la función de verosimilitud completa del parámetro θ . Por ello es recomendable presentar varios intervalos de verosimilitud. Su interpretación es muy importante, ya que un intervalo para θ , a un nivel c , indica que todo valor de θ incluido en ese intervalo, posee una verosimilitud relativa mayor o igual que c . Así, para un nivel c , valores de θ dentro de este intervalo hacen más probable a la muestra observada, que aquellos valores fuera del mismo.

Existe una diferencia entre los intervalos de verosimilitud de nivel $c \times 100\%$ y los intervalos de confianza de nivel $(1 - \alpha) \times 100\%$, ya que en estos últimos la probabilidad asignada $(1 - \alpha)$ describe, para muestras repetidas, el comportamiento del intervalo completo. En estas repeticiones, aproximadamente, $(1 - \alpha) \times 100\%$ de ellas incluirán al verdadero valor del parámetro θ_0 , de la distribución de donde provienen los datos observados. No todo intervalo de verosimilitud es necesariamente también un intervalo de confianza, pero es posible establecer una cota inferior para las coberturas de dicho intervalo. Con el fin de comprender su diferencia se define el concepto de probabilidad de cobertura.

Definición 1.8 Probabilidad de cobertura. *La probabilidad de cobertura $CP(\theta_0)$ de un intervalo aleatorio $[A(y_1, \dots, y_n), B(y_1, \dots, y_n)]$ obtenido a partir de la muestra observada (y_1, \dots, y_n) , se denota como $CP(\theta_0)$, donde*

$$CP(\theta_0) = P(A(y_1, \dots, y_n) \leq \theta_0 \leq B(y_1, \dots, y_n) | \theta = \theta_0), \quad (1.4.1)$$

y es igual a la probabilidad de que dicho intervalo aleatorio incluya al verdadero valor θ_0 del parámetro.

Un intervalo aleatorio $[A(y_1, \dots, y_n), B(y_1, \dots, y_n)]$ será llamado un *intervalo de confianza* para θ , si su probabilidad de cobertura calculada como en (1.4.1) es la misma para todos los valores θ_0 , del parámetro. Kalbfleisch (1985, p. 114) muestra que para el caso binomial, con muestras pequeñas, la probabilidad de cobertura depende del valor del parámetro θ_0 ; esto es, en ese caso los intervalos de verosimilitud no son intervalos de confianza. El nivel c de verosimilitud, con el cual se obtienen los intervalos de verosimilitud, es en cierta manera arbitrario. Comúnmente se utilizan valores de c iguales a 0.036, 0.147 y 0.258, por la relación que guardan estos valores con niveles de confianza asociados a 0.99, 0.95 y 0.90, respectivamente, cuando la función de verosimilitud es simétrica y acampanada alrededor del emv.

Ahora, en un intervalo de confianza se asigna una probabilidad a todo el intervalo, sin distinguir un valor de otro dentro del mismo. Dicha probabilidad refleja la posibilidad que el intervalo aleatorio incluya al verdadero valor del parámetro, suponiendo que se obtuvo al azar de entre una población de intervalos donde el $(1 - \alpha)\%$ de ellos lo incluyen. De esta manera, los intervalos de confianza complementan la información que proporcionan los intervalos de verosimilitud, en cuanto

a la probabilidad de que, en muestras repetidas, estos contengan el verdadero valor del parámetro θ . El intervalo de verosimilitud de nivel c es único; en contraste hay una infinidad de intervalos de confianza de nivel $(1 - \alpha)$ y solamente uno de ellos será también de verosimilitud.

La manera más común para construir intervalos de confianza es mediante el uso de cantidades pivotaes. Esta técnica data de Fisher (1945) y es un método que permite cuantificar la incertidumbre de que el intervalo de estimación incluya al verdadero valor del parámetro, que generó a la muestra observada. Enseguida se proporciona esta definición.

Definición 1.9 Cantidad pivotal. Sea Y_1, \dots, Y_n una muestra aleatoria de una distribución con parámetro θ . Si $U(Y_1, \dots, Y_n; \theta)$ es una variable aleatoria cuya distribución no depende de θ , entonces a U se le llama una cantidad pivotal para θ .

Así, para valores reales a y b , con $a < b$ es posible calcular la probabilidad

$$P[a \leq U(Y_1, \dots, Y_n; \theta) \leq b], \quad (1.4.2)$$

pues la distribución de U es independiente de θ . Si en particular, la cantidad pivotal U es lineal en θ , se podrá, a partir de U , obtener un intervalo aleatorio de estimación de la forma $[A(Y_1, \dots, Y_n), B(Y_1, \dots, Y_n)]$ para θ ; éste se consigue despejando θ en el argumento de la probabilidad mostrada en (1.4.2), esto es, puede calcularse

$$1 - \alpha = P[a \leq U(\mathbf{y}; \theta) \leq b] = P[A(Y_1, \dots, Y_n) \leq \theta \leq B(Y_1, \dots, Y_n)]. \quad (1.4.3)$$

En esta expresión, lo aleatorio es la cantidad pivotal U y los extremos aleatorios del intervalo $A(Y_1, \dots, Y_n)$ y $B(Y_1, \dots, Y_n)$, que tomarán ciertos valores particulares, una vez que la muestra ha sido observada. Debe aclararse que en un intervalo obtenido para una muestra particular (y_1, \dots, y_n) y generado con una probabilidad $1 - \alpha$, de incluir al verdadero valor del parámetro θ , no se sabrá si en realidad éste incluye o no a ese parámetro.

A manera de ejemplo, considérese un conjunto de variables aleatorias Y_1, \dots, Y_n independientes e idénticamente distribuidas como normales, con media μ y varianza σ^2 . Un ejemplo de cantidad pivotal que se utiliza para estimar la media de una distribución normal, cuando se desconoce la varianza, es la propuesta por William S. Gosset en 1909, que sigue una distribución *t-Student* con $n - 1$ grados de libertad y que se calcula como

$$U(Y_1, \dots, Y_n; \mu) = \frac{\sqrt{n-1}(\bar{Y} - \mu)}{\sqrt{\sum (Y_i - \hat{\mu})^2 / n}}, \quad (1.4.4)$$

donde $\hat{\mu} = \sum Y_i / n$ y $\hat{\sigma}^2 = \sum (Y_i - \hat{\mu})^2 / n$ son los emv de μ y de σ .

Esta cantidad pivotal es muy conocida, sin embargo, en ocasiones es difícil encontrar una cantidad pivotal o bien conocer la distribución exacta que tiene ésta.

Existen cantidades pivotaes que se llaman asintóticas porque la distribución de ellas se conoce para tamaños de muestra suficientemente grandes. En los casos en que sea posible utilizar algún resultado asintótico para aproximar dicha distribución, lo que se obtiene son intervalos de estimación aproximados para θ .

Hay dos cantidades pivotaes asintóticas importantes que es conveniente presentar. La primera es una variante de la estadística de Wald (Pawitan, 2007, p. 257) y se calcula como

$$u_\theta = \sqrt{I_{\hat{\theta}}} (\hat{\theta} - \theta). \quad (1.4.5)$$

Esta cantidad aparece en la expansión en serie de Taylor de la log-verosimilitud relativa $r(\theta; y)$ alrededor de $\hat{\theta}$, presentada en la Sección 1.3.3 y converge, en distribución, a una normal estándar. Este resultado sigue siendo válido incluso cuando θ es un parámetro de interés unidimensional que se está estimando en ausencia de información de parámetros de estorbo restantes (Sprott, 2000, p. 78). Obsérvese que u_θ depende de la muestra a través del emv y de la información observada de Fisher. Cuando la función de verosimilitud es simétrica y acampanada, bastan con estas dos cantidades para resumir la información de toda esta función.

Una segunda cantidad pivotal asintótica, que en particular se estará utilizando de manera recurrente en esta tesis, se basa en la razón de verosimilitudes. Para una muestra observada $\mathbf{y} = (y_1, \dots, y_n)$ de las variables aleatorias Y_1, \dots, Y_n con distribución $F_Y(y; \theta)$ y $\theta \in \mathbb{R}^k$ desconocido, esta cantidad pivotal conocida como estadístico de razón de verosimilitudes y que se debe a Wilks (1938), se define como

$$\Lambda_n = -2 \log \frac{L(\theta; \mathbf{y})}{L(\hat{\theta}; \mathbf{y})}. \quad (1.4.6)$$

La razón de trabajar con el doble del negativo de la log-verosimilitud, es porque resulta más conveniente para probar resultados asintóticos. El comportamiento asintótico de Λ_n es el de una distribución Ji-cuadrada con k grados de libertad, como se enuncia en el Teorema A.2, que se presenta en el Apéndice A y cuya demostración puede verse en Serfling (2002, p. 155) y en Pawitan (2007, p. 266). Así, en el caso Weibull donde $\theta = (\mu, \sigma, \beta) \in \mathbb{R}^3$, la distribución de Λ_n será la de una Ji-cuadrada con tres grados de libertad.

Cuando se tienen parámetros de estorbo, esto es, si $\theta = (\theta_1, \theta_2) \in \mathbb{R}^k$, donde $\theta_1 \in \mathbb{R}^q$ es el parámetro de interés y $\theta_2 \in \mathbb{R}^s$, es el parámetro de estorbo, con $k = q + s$, se puede utilizar la verosimilitud perfil para estimar el parámetro de interés y calcular también el estadístico de la razón de verosimilitudes. En tal caso, este estadístico seguirá una distribución Ji-cuadrada con $k - s$ grados de libertad. Por ejemplo, si cierta densidad, para la cual se cumplen las condiciones de regularidad, tiene parámetro $\theta = (\mu, \sigma, \beta) \in \mathbb{R}^3$, el parámetro de interés es $\mu \in \mathbb{R}$ y $(\sigma, \beta) \in \mathbb{R}^2$ son parámetros de estorbo, entonces $\Lambda_n \rightarrow \chi_{3-2}^2 = \chi_1^2$.

La convergencia del estadístico de la razón de verosimilitudes puede ocurrir aún en muestras pequeñas. En contraste, la cantidad pivotal u_θ requiere que la parametrización en θ , logre que la verosimilitud sea simétrica alrededor del estimador máximo verosímil. El poder obtener verosimilitudes aproximadamente simétricas requiere, en ocasiones, efectuar reparametrizaciones que se estudian con mayor detalle en la siguiente sección.

1.5 La importancia de reparametrizaciones adecuadas

En ocasiones es conveniente trabajar con un modelo reparametrizado, esto es, buscar una función uno a uno, $\varphi = g(\theta)$ del parámetro θ en el modelo estadístico considerado, tal que la función de verosimilitud reparametrizada en φ , sea más simétrica que la verosimilitud original de θ y, que la vecindad del máximo (o mínimo) pueda ser aproximada por una función cuadrática. Reparametrizaciones de este tipo permiten que sea posible aplicar resultados asintóticos, aún en muestras moderadas o pequeñas y facilita el poder dar intervalos de verosimilitud-confianza para φ y por ende para θ , gracias a la propiedad de invarianza de la verosimilitud. Otra ventaja de estas reparametrizaciones es al utilizar procesos numéricos de optimización, ya que la búsqueda del máximo (o mínimo) es más eficiente Ross (1970).

En la Sección 1.3.3 se expuso que el término $F_3(\hat{\theta}; \mathbf{x})$ de la expansión (1.3.11), proporciona información acerca de la simetría de la función de verosimilitud, alrededor del máximo verosímil. Si θ es unidimensional y si se logra obtener una función ϕ tal que $F_3(\hat{\phi}; \mathbf{y}) \approx 0$, entonces la verosimilitud del nuevo parámetro ϕ será más simétrica que la verosimilitud original de θ . Para mostrar lo anterior consideremos el ejemplo siguiente.

Ejemplo 1.4 *Considérese una muestra observada (y_1, \dots, y_n) de Y_1, \dots, Y_n variables aleatorias independientes e idénticamente distribuidas como exponenciales, con función de densidad*

$$f(y; \theta) = \frac{1}{\theta} \exp(-y/\theta) I_{[0, \infty)}(y), \text{ con } \theta > 0.$$

La muestra es tal que $(y_1 + y_2) = 3$. El estimador de máxima verosimilitud de θ es $\hat{\theta} = \bar{y} = 1.5$ y la función de verosimilitud relativa para θ está dada por

$$R(\theta; \mathbf{y}) = \left(\frac{\hat{\theta}}{\theta} \right)^n \exp \left[n \left(1 - \frac{\hat{\theta}}{\theta} \right) \right] = \left(\frac{1.5}{\theta} \right)^2 \exp \left[2 \left(1 - \frac{1.5}{\theta} \right) \right]. \quad (1.5.1)$$

Sprott y Viveros (1984), muestran que la gráfica de esta verosimilitud relativa se logra simetrizar aún cuando se cuente con tan solo dos observaciones, si se hace uso de la transformación $\phi = \theta^{-1/3}$, esto es, $\phi^{-3} = \theta$.

La verosimilitud relativa del nuevo parámetro ϕ resulta

$$R(\phi; \mathbf{y}) = \left(\frac{\phi}{\widehat{\phi}}\right)^{3n} \exp \left\{ n \left[1 - \left(\frac{\widehat{\phi}}{\phi}\right)^3 \right] \right\} = \left(\frac{\phi}{\widehat{\phi}}\right)^6 \exp \left\{ n \left[1 - \left(\frac{\widehat{\phi}}{\phi}\right)^3 \right] \right\}. \quad (1.5.2)$$

La Figura 1.4(a) muestra la verosimilitud relativa del parámetro θ , en la cual se observa un marcado sesgo a la derecha, mientras que en la Figura 1.4(b) se puede observar que la verosimilitud del parámetro ϕ es simétrica alrededor del emv $\phi = 0.874$. La transformación no es fortuita, ya que se obtuvo buscando la reparametrización

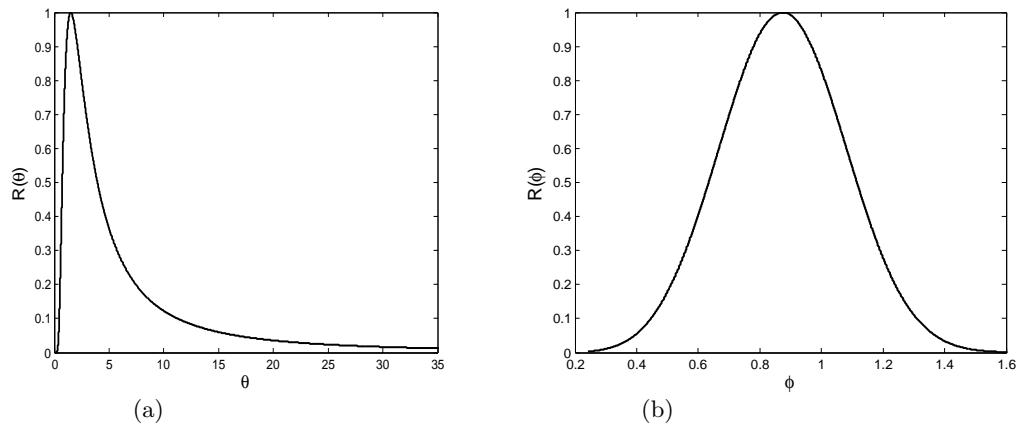


Figura 1.4: Ejemplo 1.3: Datos exponenciales (a) Verosimilitud relativa de θ (b) Verosimilitud relativa de ϕ .

potencia $\phi = \theta^k$ y una $k \in \mathbb{R}$ que resulte de resolver $F_3(\phi; \mathbf{y}) = 0$. El valor de k que satisface la ecuación es $k = -1/3$.

Con lo anterior se muestra que, aún contando con tan sólo dos observaciones, es posible encontrar una transformación adecuada con la cual se obtenga una buena aproximación a una normal y se satisfagan resultados asintóticos.

1.6 La robusticidad en la estimación

Cuando se plantea un modelo estadístico para describir un fenómeno aleatorio de interés siempre debe tenerse en cuenta que el modelo propuesto es tan solo una aproximación a la realidad compleja que intenta describirse. Podría darse el caso que existan varios modelos estadísticos que sean razonables para describir el fenómeno de interés. Por tanto, es importante evaluar qué tan sensibles son las inferencias sobre los parámetros de interés (que sean comunes a los modelos considerados, como por ejemplo los cuantiles) y las conclusiones a las que se llega al considerar varios modelos posibles. Si las inferencias no cambian mucho y son estables, se dirá que se trata de un ejemplo o caso *robusto*. En cambio, si pequeños cambios en el modelo implican cambios fuertes en las inferencias o en las conclusiones, se dirá que se tiene

un ejemplo no robusto. En tal caso se requiere obtener mayor información sobre el problema o aumentar el tamaño de la muestra para poder efectuar inferencias válidas sobre los parámetros de interés.

La función de verosimilitud perfil del parámetro de interés permite evaluar fácilmente si se cuenta con un ejemplo robusto o no, al graficarla para cada uno de los modelos considerados. Sprott y Farewell (1993) presentan un ejemplo robusto y otro que no lo es, en el contexto de las inferencias sobre la diferencia $\delta = \mu_2 - \mu_1$ entre las medias μ_1 y μ_2 , de dos poblaciones normales, donde σ_1 y σ_2 son las desviaciones estándar respectivas y $\rho = \sigma_2/\sigma_1$ denota su razón. Este ejemplo estudia la sensibilidad de las inferencias sobre δ frente a cambios en ρ . Para estimar δ se utiliza generalmente una cantidad pivotal que tiene distribución t de Student y la cual supone que ρ es conocida e igual a uno. Para justificar esta afirmación tan fuerte, generalmente se realiza una prueba de hipótesis sobre la igualdad de varianzas, y en el caso de no rechazarla, suele considerarse válido afirmar que efectivamente $\rho = 1$.

Sprott y Farewell muestran que es mejor dar un intervalo de estimación para ρ basado en las dos muestras observadas. Posteriormente consideran varios valores de ρ en este intervalo y proceden a graficar la verosimilitud relativa perfil condicional de δ para estos valores de ρ . Se debe observar si las curvas obtenidas son cercanas y parecidas entre sí. En el caso de tener anchos o aperturas muy distintas con respecto a su máximo, indicarán que la precisión en las estimaciones de δ cambian significativamente al cambiar ρ . En estas curvas debe también observarse si el valor de verosimilitud asignado a $\delta = 0$ es muy diferente o no, pues este valor es equivalente a la igualdad de las medias normales. Cuando hay cambios importantes, se dice que se tiene un caso no robusto, pues pequeños cambios en el valor de ρ , que estarían sustentados por los datos observados, conducen a conclusiones muy distintas sobre las medias.

Para ejemplificar estas ideas Sprott y Farewell presentan, como caso robusto, un ejemplo propuesto en Kalbfleisch (1985, p. 214), donde se comparan las distribuciones del tamaño de los huevos que las madres Cuco dejan en nidos de aves de dos especies distintas (Reyezuelos y Currucas). Se supone que estos tamaños siguen una distribución normal e interesa hacer inferencia sobre δ , la diferencia de las medias; en particular interesa la plausibilidad asignada al valor $\delta = 0$. La Figura 1.5(a) muestra la verosimilitud perfil relativa de ρ ; en ella está marcado el intervalo del 15% de verosimilitud (0.64, 2.1), el cual incluye al valor $\rho = 1$. La Figura 1.5(b) muestra las verosimilitudes perfiles de δ condicionadas a cuando ρ toma como valor los extremos de este intervalo, así como la estimación de máxima verosimilitud, que fue de 1.19. Se puede notar que las tres curvas están cercanas entre sí y que además le asignan muy poca plausibilidad al valor $\delta = 0$, por lo que hay evidencia fuerte de que las medias μ_1 y μ_2 no son iguales. Por otra parte, el ejemplo no robusto que Sprott y Farewell presentan, compara el logaritmo de los tiempos de falla de dos tipos de engranes de plástico que se prueban a dos temperaturas distintas (Kalbfleisch, 1985, p. 215). Se tienen dos muestras pequeñas de tamaños 4 y 8, y se realiza un análisis

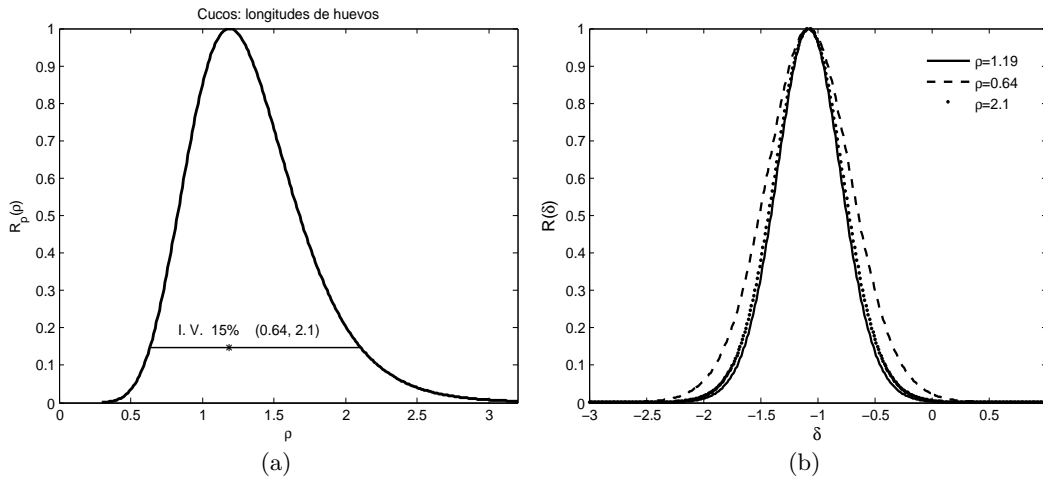


Figura 1.5: Longitud de huevos de Cucos: (a) Perfil de ρ , la raíz del cociente de varianzas y (b) Perfiles condicionadas de δ para tres valores de ρ .

similar. La Figura 1.6(a) muestra la verosimilitud relativa perfil de ρ , donde se marca el intervalo del 15% de verosimilitud $(0.49, 2.95)$, así como el emv $\hat{\rho} = 1.1$. Este intervalo nuevamente incluye el valor de $\rho = 1$. La Figura 1.6(b) muestra cómo cambia notoriamente la amplitud de las perfiles relativas de δ condicionadas en ρ , cuando éste toma el valor 1.1 y los valores de los extremos del intervalo mencionado. No solamente cambia la forma y anchura de estas curvas sino también la verosimilitud que asignan al valor de $\delta = 0$ es muy distinta. Para el valor $\rho = 2.95$, no hay evidencia en contra de la igualdad de medias. En cambio, para $\rho = 1.1$ o $\rho = 0.49$, los datos sustentan fuertemente que una media es mayor que la otra. Este es un ejemplo no robusto, pues las conclusiones sobre la igualdad de las medias cambian dramáticamente al considerar diferentes valores razonables que ρ puede tomar, a la luz de los datos.

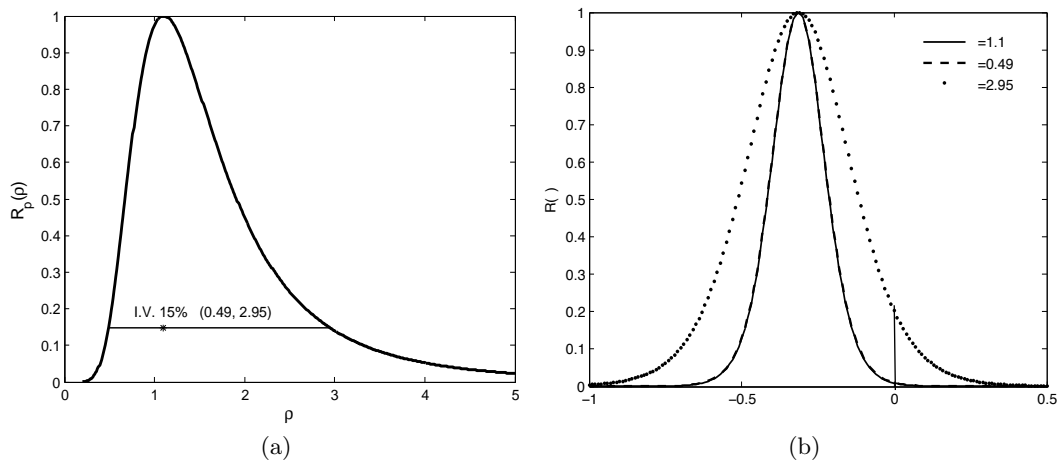


Figura 1.6: Engranajes de plástico: (a) Perfil de ρ , la raíz del cociente de varianzas y (b) Perfiles condicionadas de δ para tres valores de ρ .

Otro ejemplo no robusto, también para un caso donde se desean comparar dos medias poblacionales, lo presenta Sprott (2000, p. 141) al analizar los datos pareados de Darwin, referentes a parejas de plantas sembradas en macetas distintas, donde una de ellas fue autofertilizada y en la otra se utilizó fertilización cruzada. En este problema interesa comparar la diferencia entre las medias de las alturas de dichas plantas. Es un ejemplo no robusto, pues dependiendo del modelo supuesto para las alturas de las plantas en cada población, se llega a una conclusión distinta sobre la igualdad de las medias. Cuando se considera que la distribución madre es normal, se concluye que las medias podrían ser iguales. En contraste, si se considera por ejemplo una distribución Cauchy, se concluye que las medias son diferentes.

Es difícil identificar qué condiciones están asociadas a la ocurrencia de un ejemplo no robusto. A través de simulaciones efectuadas para el caso de muestras de dos poblaciones normales, donde se desea estimar la diferencia de medias, se observó que los ejemplos no robustos se presentan generalmente cuando los tamaños de muestras son pequeños, marcadamente desiguales y además la razón de desviaciones estándar es grande, $\rho > 4$. En algunos otros casos suelen estar asociados a muestras pequeñas con alguna otra característica que las hace poco informativas.

Cuando se tiene un ejemplo robusto existe la seguridad de que las inferencias son estables y por lo tanto más confiables, pues a pesar de la incertidumbre intrínseca en el modelo, se llegará siempre a las mismas conclusiones sobre los parámetros de interés. En contraste, cuando los ejemplos son no robustos, no se pueden obtener dichas conclusiones y para poder afirmar algo sobre el problema de interés se debe recabar mayor información.

En muchos problemas se ha observado que los cuantiles son cantidades muy informativas sobre la distribución del fenómeno aleatorio de interés y en muchos casos es posible reparametrizar el modelo probabilístico de interés en términos de un cuantil dado. Por ejemplo, para las distribuciones que pertenecen a la familia de localización y escala, siempre es posible reparametrizar en términos de un cuantil; igualmente para las distribuciones de valores extremos y muchas otras más. Como se mostrará en la siguiente sección, estos parámetros son muy informativos ya que las inferencias sobre ellos suelen ser generalmente robustas frente a cambios importantes en los modelos considerados. Ejemplos de ello se mostrarán en el Capítulo 4 de esta tesis.

1.7 Los cuantiles como parámetros estables

En muchos contextos se requiere comparar diversos modelos probabilísticos que son candidatos a describir un fenómeno aleatorio de interés. La comparación entre modelos distintos no siempre es inmediata o fácil, sobre todo si poseen diferentes parámetros. Sin embargo, existen parámetros que son comunes a todos los modelos, son sencillos de interpretar y permiten evaluar la proximidad entre diferentes

distribuciones; estos parámetros son los cuantiles. Los cuantiles de una distribución pueden llegar a ser incluso de mayor interés que los parámetros originales de la distribución, debido a la interpretación que tienen en términos de probabilidades acumuladas y por ser invariantes bajo escala (Sprott, 2000, p. 122). A continuación se presenta la definición de un cuantil de probabilidad α de una distribución.

Definición 1.10 *Cuantil de probabilidad α de una distribución.* Sea Y una variable aleatoria con función de distribución $F_Y(y; \theta)$. Sea $0 < \alpha < 1$. La cantidad Q_α se define como el cuantil de probabilidad α de F_Y , si éste es el valor más pequeño que satisface la relación $F_Y(Q_\alpha; \theta) \geq \alpha$.

Cuando F_Y es continua y estrictamente creciente, el cuantil Q_α suele definirse como el valor que toma la variable aleatoria Y , y que satisface la igualdad $F_Y(Q_\alpha) = \alpha$. Como en este caso F_Y es invertible, entonces el cuantil se obtiene simplemente aplicando la función de distribución inversa a la probabilidad α , es decir, $Q_\alpha = F_Y^{-1}(\alpha)$. Así para variables aleatorias absolutamente continuas un cuantil de probabilidad α se definirá como la función de distribución inversa correspondiente, aplicada a dicha probabilidad.

En el caso de variables aleatorias discretas, como las funciones de distribución asociadas son escalonadas, es usual definir a los cuantiles a través de la función inversa generalizada de la distribución correspondiente, definición que aquí no se presentará, pues en esta tesis se considerarán exclusivamente cuantiles de variables aleatorias absolutamente continuas.

El reparametrizar un modelo estadístico en términos de un cuantil presenta muchas ventajas. La principal es que, una función de verosimilitud en términos de un cuantil suele ser mucho más simétrica que la que se obtiene con otro tipo de parámetro. Esto facilita el proceso de maximización numérica pues resulta más eficiente y rápido encontrar al emv y también las verosimilitudes perfiles de los parámetros de interés. Sin embargo, la principal ganancia al poder obtener una verosimilitud más simétrica, es que los resultados asintóticos de máxima verosimilitud serán aplicables al tamaño de muestra que se esté considerando. Así, se facilitará proporcionar intervalos óptimos de estimación para los parámetros de interés, como son los intervalos de verosimilitud-confianza.

Otra ventaja de parametrizar en términos de cuantiles radica en la facilidad con la que se pueden dar valores iniciales en un proceso de optimización. La recomendación es utilizar el cuantil empírico asociado y con ello la convergencia al óptimo se reduce a unas cuantas iteraciones. Una ventaja más es que la matriz de información observada de Fisher suele ser fácilmente invertible y no-singular, a diferencia de cuando se utilizan otras parametrizaciones del modelo. Por otra parte, una gran ventaja que presentan los cuantiles es que, en general, se ven poco afectados por cambios pequeños en los parámetros restantes del modelo, así como por observaciones extremas discrepantes en la muestra, por lo que son considerados parámetros

robustos. Cuando un parámetro tiene estas características, Ross (1970) lo define como un parámetro estable.

Hay que resaltar que para una muestra de n observaciones, el cuantil empírico más pequeño que se puede estimar con los datos es el asociado a una probabilidad de $1/(n+1)$, mientras que el más grande está asociado a la probabilidad $n/(n+1)$. La verosimilitud perfil de la mediana Q_{50} suele, por lo general, ser simétrica, aunque el tamaño de muestra considerado es importante, como puede verse en la Figura 1.7 donde se comparan las verosimilitudes relativas del parámetro θ de una distribución exponencial y de la mediana Q_{50} , para muestras de tamaños $n = 2$ y $n = 10$. Por

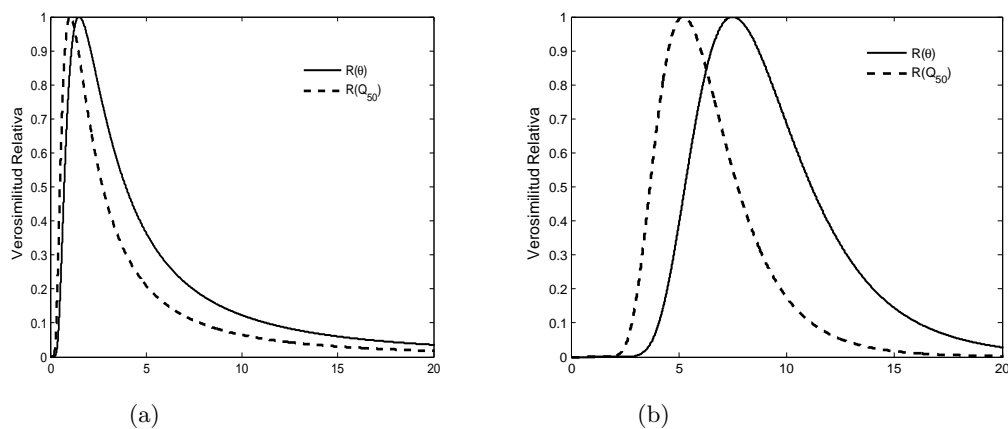


Figura 1.7: Datos exponenciales: Verosimilitud relativa de θ (—) y de Q_{50} (- -) para (a) $n = 2$ y (b) $n = 10$.

otra parte, la verosimilitud de un cuantil que se encuentre en las colas de la distribución, como $Q_{0.05}$, suele tener una verosimilitud perfil asimétrica (Sprott, 2000, p. 184) indicando la falta de información que sobre éste, tienen los datos observados. Así, para poder estimar bien cuantiles pequeños o grandes, se requiere contar con muestras mucho más grandes. En algunos ejemplos del Capítulo 4 se mostrará que la verosimilitud perfil de cuantiles suele ser mucho más simétrica que la perfil de un parámetro umbral.

Los cuantiles resultan ser muy útiles para evaluar la cercanía entre distribuciones distintas, en particular cuando se tiene una familia de distribuciones que está cercana a un modelo límite más sencillo. Tal es el caso de la familia Weibull de tres parámetros, la cual tiene como modelo límite a la distribución Gumbel (Cheng y Iles, 1990), que pertenece a las distribuciones de localización y escala. En el Capítulo 4 se verá que el modelo Weibull describe bien el comportamiento aleatorio de muchos de los ejemplos prácticos que se presentan.

A manera de ejemplo, se muestra aquí cuál sería Q_α , el cuantil de probabilidad

α para el modelo Weibull de mínimos, cuya función de distribución es

$$F(x; \mu, \sigma, \beta) = \left\{ 1 - \exp \left[- \left(\frac{x - \mu}{\sigma} \right)^\beta \right] \right\} I_{[\mu, \infty)}(x), \quad (1.7.1)$$

y donde $\mu \in \mathbb{R}$, $\sigma > 0$ y $\beta > 0$, son los parámetros de localización, escala y forma, respectivamente. El cuantil Q_α de este modelo, se obtiene al aplicar la distribución inversa F^{-1} a la probabilidad α y resulta ser una función de los tres parámetros del modelo, así

$$Q_\alpha = F^{-1}(\alpha) = \mu + \sigma [-\log(1 - \alpha)]^{\frac{1}{\beta}}. \quad (1.7.2)$$

El modelo Weibull se puede reparametrizar también como una distribución propuesta por Von Mises (1954) y de manera independiente por Jenkinson (1955), conocida como Distribución Generalizada de Valores Extremos o DGVE, cuya distribución para mínimos está dada por

$$F(x; a, b, c) = \begin{cases} 1 - \exp \left[- \left(1 - c \left(\frac{x+a}{b} \right) \right)^{-\frac{1}{c}} \right] I_{A(a,b,c)}(x), & \text{si } c \neq 0, \\ \{ 1 - \exp \left[- \exp \left(\frac{x+a}{b} \right) \right] \} I_{\mathbb{R}}(x), & \text{si } c = 0, \end{cases} \quad (1.7.3)$$

donde $A(a, b, c) = [-a + (b/c), \infty)$ si el parámetro de forma $c < 0$, lo cual corresponde a la familia de distribuciones Weibull y $A(a, b, c) = (-\infty, -a + (b/c)]$ si $c > 0$, que corresponde a la familia de distribuciones Fréchet; en el límite, cuando $c = 0$ se tiene una distribución Gumbel.

Los parámetros a, b, c de esta distribución tienen también la característica de ser muy estables y la función de verosimilitud asociada en términos de (a, b, c) , es mucho más simétrica que la verosimilitud obtenida para los parámetros de una familia Weibull o de una Fréchet.

La relación que guardan los parámetros de la distribución Weibull de mínimos con la DGVE es

$$\mu = -a - b/|c|, \quad \sigma = b/|c|, \quad \beta = 1/|c|, \quad (1.7.4)$$

y los cuantiles para la DGVE son

$$Q_\alpha = \begin{cases} -a + b \log [-\log(1 - \alpha)], & \text{si } c = 0, \\ -a + (b/c) \{ 1 - [-\log(1 - \alpha)]^{-c} \}, & \text{si } c \neq 0. \end{cases} \quad (1.7.5)$$

En la Tabla 1.1 se muestran cuatro distribuciones muy cercanas entre sí, tomando como base la parametrización de la DGVE. Los valores de a y b son iguales en los cuatro modelos; solamente cambia, ligeramente, el valor del parámetro de forma c . Se puede observar que los cuantiles de probabilidades de 0.05 a 0.95 son muy parecidos en dichos modelos y, difieren un poco en los cuantiles de probabilidades 0.01 y 0.99. En contraste, si los primeros tres modelos se hubiesen parametrizado en términos de los parámetros de la familia Weibull, esto es, en términos de (μ, σ, β) , se notaría que

Weibull			DGVE			Algunos Cuantiles bajo el modelo Weibull				
Umbral	Forma	Escala	Localización	Escala	Forma	$Q_{0.01}$	$Q_{0.05}$	$Q_{0.5}$	$Q_{0.95}$	$Q_{0.99}$
μ	β	σ	a	b	c					
-77	25	75	2	3	-0.04	-14.605	-10.402	-3.092	1.362	2.724
-302	100	300	2	3	-0.01	-15.488	-10.780	-3.098	1.310	2.617
-377	125	375	2	3	-0.008	-15.550	-10.806	-3.098	1.306	2.610
Gumbel			2	3	0	-15.800	-10.911	-3.099	1.292	2.581

Tabla 1.1: Parámetros y cuantiles de tres modelos Weibull (μ, σ, β) cercanos al modelo límite Gumbel correspondiente. Reparametrización en términos de los parámetros de la DGVE (a, b, c) y algunos cuantiles.

dichos parámetros son sumamente distintos en magnitud; sin embargo las distribuciones siguen siendo sumamente cercanas y los parámetros de una Weibull dificultan el percatarse de ello. Nótese entonces, cómo los cuantiles de estos modelos permiten compararlos y evaluar su cercanía. Se tiene así, un ejemplo con tres modelos Weibull que parecen muy distintos entre sí, pero en realidad, hay mucha cercanía entre ellos mismos y su modelo límite Gumbel.

Algunos autores recomiendan utilizar los cuantiles por la estabilidad que se observa en sus estimaciones puntuales. Sin embargo, una estimación puntual no es lo más importante. Debe considerarse, en cambio, la capacidad que en general tienen los cuantiles para simetrizar la superficie de verosimilitud. Gracias a ello pueden aplicarse resultados asintóticos para muestras incluso pequeñas y así estimar, de manera óptima, los parámetros de interés en términos de intervalos de verosimilitud-confianza. Más aún, otro resultado importante que se enfatiza aquí y aparentemente no se ha hecho antes, es la robusticidad que presentan los intervalos de verosimilitud perfil de los cuantiles, frente a cambios en los modelos considerados (siempre y cuando éstos sean razonables para la muestra observada). Estas propiedades se verificarán con las simulaciones realizadas que se muestran en el Capítulo 3, donde también se muestra que los intervalos para los cuantiles considerados presentan buenas coberturas, inclusive en el caso de muestras pequeñas.

Capítulo 2

Críticas injustificadas a la función de verosimilitud

En la literatura estadística se han presentado muchas críticas a la función de verosimilitud, que son injustificadas en su mayoría, (Smith y Weissman (1985), Smith y Naylor (1987b), Green et al. (1994), Hirose y Lai (1997), etcétera). Estas críticas surgen de utilizar una definición incorrecta de la verosimilitud, o de interpretarla equivocadamente. La mayoría de estas críticas se originan porque se obtienen verosimilitudes no acotadas y por tanto no existe el emv. Otro tipo de críticas son respecto a la forma aplanada que en algunas ocasiones toma la verosimilitud perfil de un parámetro umbral, por lo que ésta se califica como no informativa, con un emv muy inestable y poco robusto.

En este capítulo se analizará a detalle el caso de verosimilitudes no acotadas que son resultado de utilizar densidades que poseen un parámetro umbral y una singularidad. El problema, básicamente, se debe a utilizar la aproximación continua a la verosimilitud, cuando no se cumplen las condiciones de regularidad que deben darse para que esta aproximación sea válida. Como resultado de ello, la singularidad de la función de densidad se hereda a la función de verosimilitud y con ello existe una trayectoria en el espacio parametral, donde la verosimilitud tiende a infinito. Esto de ninguna manera significa que esos valores de los parámetros hagan más probable a la muestra observada (Meeker y Escobar, 1998, p. 275). En este tipo de situaciones puede ocurrir que la función de verosimilitud perfil tenga forma de J . En tal caso no existirá un máximo local, sino solamente el punto donde se alcanza la singularidad. En la Sección 2.5 se discutirán estas situaciones, con mayor detalle, para distribuciones de tres parámetros, como son la Weibull y lognormal.

Con respecto a las verosimilitudes perfiles obtenidas con L_C , y que tienen forma aplanada, interesarán solamente aquellas que provienen de densidades que poseen un parámetro umbral y singularidades. Generalmente alcanzan un máximo local que está muy alejado de la singularidad, por lo que son muy asimétricas y planas. Esta asimetría marcada se puede deber a varias condiciones; entre ellas está el contar con una muestra pequeña, o que el modelo considerado se encuentre cercano a un modelo límite más sencillo. Algunos autores (Smith y Naylor, 1987b y Green et al., 1994), critican fuertemente las verosimilitudes perfiles planas, pues centran su atención en una estimación puntual y no consideran toda la información que la función de verosimilitud completa proporciona a través de intervalos de estimación.

La función de verosimilitud restringida que se presentó en el Capítulo 1, o bien, la función de verosimilitud discretizada, que se propone en esta tesis y se presenta en el Capítulo 3, solucionan el problema de las verosimilitudes no-acotadas, debidas a la existencia de singularidades. El caso de verosimilitudes planas generalmente se relaciona con un problema de selección de modelos. Aún utilizando la verosimilitud restringida e incluso la verosimilitud discretizada se seguirán teniendo estas formas. Por ello, en la Sección 2.4 se muestra que algunas distribuciones de tres parámetros poseen modelos límites más simples y regulares que pueden describir bien al fenómeno aleatorio de interés.

2.1 Verosimilitudes perfiles no-acotadas

En esta sección se muestran algunos ejemplos en los que se obtiene una verosimilitud perfil no-acotada, que resulta del uso indebido de la aproximación continua a la función de verosimilitud. El problema se soluciona cuando se considera que todo instrumento de medición tiene precisión finita y que las verosimilitudes son proporcionales a probabilidades. Con ello la función de verosimilitud resulta ser siempre acotada.

En el primer ejemplo se utiliza una mezcla de dos distribuciones normales para modelar cierta muestra y en el segundo se presenta el uso de densidades que poseen un parámetro umbral y una singularidad. Para ambos casos, en la literatura estadística se ha reportado que el emv no existe.

Ejemplo 2.1 Sean Y_1, Y_2, \dots, Y_n variables aleatorias independientes e idénticamente distribuidas como una mezcla de dos normales con medias μ_1, μ_2 y varianzas σ_1^2 y σ_2^2 , respectivamente. El parámetro de proporción de mezcla es p . La densidad de mezcla de una de estas variables aleatorias es entonces,

$$\begin{aligned} g(y; \mu_1, \sigma_1, \mu_2, \sigma_2, p) &= p f_1(y; \mu_1, \sigma_1) + (1 - p) f_2(y; \mu_2, \sigma_2) \\ &= \frac{p}{\sigma_1 \sqrt{2\pi}} \exp \left[-\frac{(y - \mu_1)^2}{2\sigma_1^2} \right] + \frac{1 - p}{\sigma_2 \sqrt{2\pi}} \exp \left[-\frac{(y - \mu_2)^2}{2\sigma_2^2} \right]. \end{aligned} \quad (2.1.1)$$

Dada una muestra observada $\mathbf{y} = (y_1, y_2, \dots, y_n)$, la aproximación continua a la verosimilitud es el producto de las densidades correspondientes,

$$L(\mu_1, \sigma_1, \mu_2, \sigma_2, p; \mathbf{y}) \propto \prod_{i=1}^n [p f_1(y_i; \mu_1, \sigma_1) + (1 - p) f_2(y_i; \mu_2, \sigma_2)].$$

La log-verosimilitud es

$$l(\mu_1, \sigma_1, \mu_2, \sigma_2, p; \mathbf{y}) = \sum_{i=1}^n \log \left\{ \frac{p}{\sigma_1 \sqrt{2\pi}} \exp \left[-\frac{(y_i - \mu_1)^2}{2\sigma_1^2} \right] + \frac{1 - p}{\sigma_2 \sqrt{2\pi}} \exp \left[-\frac{(y_i - \mu_2)^2}{2\sigma_2^2} \right] \right\}. \quad (2.1.2)$$

Cuando $\mu_1 = y_i$ para $i = 1, \dots, n$ y σ_1 tiende a cero, la densidad de mezcla tiene n singularidades en $\sigma_1 = 0$. Por tanto esta verosimilitud no está acotada.

En el siguiente ejemplo se tiene también una verosimilitud perfil no-acotada, resultado de utilizar nuevamente la aproximación continua en una densidad que posee un parámetro umbral y una singularidad, como es el modelo Weibull de tres parámetros. Este modelo es muy utilizado en una gran variedad de aplicaciones ya que, como se verá en la Sección 2.5.1, esta densidad puede tomar una gran variedad de formas, dependiendo de los valores que tomen sus parámetros.

Ejemplo 2.2 Sean Y_1, Y_2, \dots, Y_n variables aleatorias independientes e idénticamente distribuidas como una Weibull de mínimos con parámetros (μ, σ, β) , donde μ es un parámetro umbral, σ es un parámetro de escala y β es un parámetro de forma. La función de densidad de una de estas variables es

$$f(y; \mu, \sigma, \beta) = \frac{\beta}{\sigma} \left(\frac{y - \mu}{\sigma} \right)^{\beta-1} \exp \left[- \left(\frac{y - \mu}{\sigma} \right)^\beta \right] I_{[\mu, \infty)}(y), \quad (2.1.3)$$

donde $\sigma > 0$ y $\beta > 0$.

Si $\mathbf{y} = (y_1, \dots, y_n)$ es una realización de una muestra aleatoria de estas variables, la aproximación continua a la función de verosimilitud se calcula como

$$\begin{aligned} L_c(\mu, \sigma, \beta; \mathbf{y}) &\propto \prod_{i=1}^n f(y_i; \mu, \sigma, \beta) \\ &= \frac{\beta^n}{\sigma^n} \left(\prod_{i=1}^n \frac{y_i - \mu}{\sigma} \right)^{\beta-1} \exp \left[- \sum \left(\frac{y_i - \mu}{\sigma} \right)^\beta \right] I_{(-\infty, y_{(1)}]}(\mu) I_{(0, \infty)}(\beta) I_{(0, \infty)}(\sigma). \end{aligned} \quad (2.1.4)$$

Se puede observar que si en (2.1.4) se tiene que $\beta < 1$, la verosimilitud tenderá a infinito cuando $\mu \rightarrow y_{(1)}$. Esto no significa que la probabilidad de la muestra observada para esa región del espacio de parámetros sea mayor que en otras regiones. El problema reside en que se está utilizando una aproximación continua a la verosimilitud cuando no se satisfacen las condiciones necesarias para hacerlo, ya que el soporte de la variable aleatoria está restringido por el parámetro umbral μ . Como resultado de ello (Berger et al., 1999) reportan que el emv no existe en este caso.

Existen muchas otras situaciones donde se pueden encontrar verosimilitudes no-acotadas, pero en esta tesis el interés reside en aquellas que se originan por utilizar densidades que poseen un parámetro umbral y una singularidad, como en el ejemplo anterior. Es por ello que a continuación se expone el papel que juega un parámetro umbral en una densidad, las formas que puede tomar la verosimilitud perfil de este parámetro cuando se utiliza la aproximación continua, así como formas alternativas para estimar este parámetro de interés.

2.2 La verosimilitud perfil de un parámetro umbral

El parámetro umbral de una distribución juega un rol muy importante en muchas aplicaciones. Sin embargo, al ser una cota superior o inferior que delimita el soporte de la variable aleatoria, ocasiona algunos problemas de estimación que se analizarán en las secciones siguientes.

Algunas distribuciones que tienen un parámetro umbral son la Weibull, la log-normal y la gamma de tres parámetros, que son utilizadas para describir una gran variedad de fenómenos aleatorios. El parámetro umbral suele ser de interés, por ejemplo, para determinar cotas inferiores de tiempos de falla. Éstas generalmente se usan como tiempos de garantía, ya que se pretende que la mayoría de los componentes fallen exclusivamente después de dicho tiempo de garantía y no antes. O bien, puede ser de interés conocer el tiempo a partir del cual es posible detectar una enfermedad. Así pues, es muy común que el objetivo principal de un estudio sea el estimar una cota superior, o bien inferior, para la variable aleatoria de interés.

Debe tenerse en cuenta que un parámetro umbral puede tomar valores negativos, a pesar de fungir como una cota inferior de variables aleatorias que describen fenómenos no negativos, como lo son mediciones de lluvias, tiempos, pesos de personas, etcétera. Es importante señalar esto pues, por lo general, en los modelos con umbral no existe la restricción matemática de que este parámetro deba ser positivo. Autores como Hirose and Lai (1997) mencionan que las estimaciones de los parámetros pueden sujetarse a la condición de que el umbral sea positivo. Sin embargo, aquí se verá que basta con pedir que la probabilidad de observar valores negativos para la variable en estudio, sea despreciable bajo el modelo estimado. Esto puede cumplirse aún con un umbral negativo, como se verá en los ejemplos del Capítulo 4.

Generalmente es difícil estimar un umbral a partir de la muestra observada, sobre todo si el umbral está alejado de la región donde, con mayor probabilidad, se observa la variable aleatoria. En ese caso la estimación del umbral equivale a realizar una extrapolación fuerte, a partir de los datos observados. Es muy frecuente que los datos no contengan mucha información acerca del umbral, lo cual se verá reflejado en la función de verosimilitud perfil correspondiente, la cual puede resultar muy asimétrica y abierta, o incluso ser de forma aplanada. Esto indica que, para este parámetro, un amplio rango de valores son igualmente sustentados por los datos.

La estimación de un parámetro umbral ha sido y es actualmente un tema de gran interés, pues se utiliza en una gran variedad de aplicaciones. Esta necesidad ha motivado varias propuestas para su estimación. Sin embargo, al ser el umbral, un parámetro que delimita el soporte de la distribución en estudio, las condiciones de regularidad clásicas de la teoría de máxima verosimilitud, no se satisfacen y por lo tanto no se cumple la teoría asintótica tradicional de máxima verosimilitud, para este parámetro. Estas condiciones de regularidad y el teorema de máxima verosimi-

litud se enuncian en las Secciones A.1 y A.2 del Apéndice A.

Para estimar un parámetro umbral, en la literatura estadística por lo general se considera la aproximación continua a la verosimilitud L_C , que se definió en la Sección 1.3.5. Smith (1985) utiliza L_C para ciertos casos no regulares y encuentra que en algunos circunstancias, el estimador de máxima verosimilitud existe y posee las mismas propiedades asintóticas que se tienen en los casos regulares. En otros casos, el emv puede existir, pero no ser asintóticamente eficiente o asintóticamente normal. También puede ocurrir que el emv no se pueda calcular como solución de las ecuaciones de verosimilitud. Ejemplo de ello es el caso de la exponencial con tiempo de vida garantizado, donde la función de verosimilitud es discontinua en la observación más pequeña y por tanto no es diferenciable en ese punto. Finalmente, el emv puede no existir como el argumento donde L_C alcance el máximo, sobre todo cuando L_C tenga singularidades y sea no acotada. En esta tesis se verá que al usar la verosimilitud restringida o la verosimilitud discretizada, que se propone en el siguiente capítulo, siempre es posible encontrar un emv y la estimación se realiza sin problema alguno.

A través de estudios de simulación que, bajo diferentes escenarios, realizaron Smith y Weissman (1985), proponen una clasificación de las diferentes formas que puede tomar la verosimilitud perfil de un parámetro umbral, cuando para su estimación se utiliza la aproximación continua a la verosimilitud L_C . Estas posibles formas se muestran en la Figura 2.1.

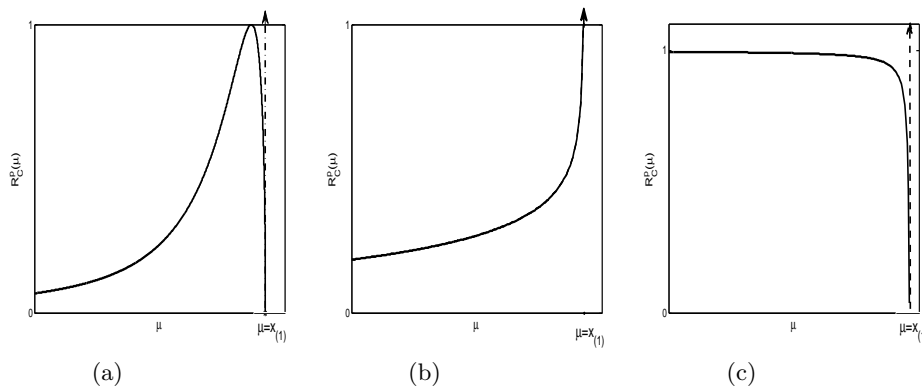


Figura 2.1: Formas de la verosimilitud perfil del umbral cuando se utiliza la aproximación continua a la función de verosimilitud.

Puede observarse que en todas ellas existe una singularidad en $\mu = x_{(1)}$, la cual se indica en las gráficas mediante una flecha punteada, ascendente. En la forma correspondiente a la Figura 2.1(a) se tiene un máximo local. La verosimilitud perfil del umbral mostrada en la Figura 2.1(b) tiene forma de J y es un caso donde no existe máximo local alguno aparte de la singularidad. En cambio, la forma presentada en la Figura 2.1(c) es aplanada y posee un máximo local, aunque muy alejado

de la singularidad. A pesar de que Smith y Weissman (1985) consideran sólo estas tres formas posibles, entre ellas existen variantes con formas intermedias que toma la verosimilitud perfil de un parámetro umbral.

En ocasiones estimar un cuantil de probabilidad pequeña puede jugar un rol similar al de un parámetro umbral, proporcionando una cota inferior para la variable de interés. Esto, en el sentido de que a la izquierda de dicho cuantil solamente ocurrirán un porcentaje pequeño de las observaciones. De hecho, un umbral es el cuantil de probabilidad cero de una distribución.

Considerando lo anterior, en esta tesis se estudia la posibilidad de estimar un cuantil pequeño (o bien grande), a manera de dar una cota inferior (o superior) para una proporción importante de los datos. Una ventaja de ello es que los cuantiles son parámetros estables (Ross, 1970), en el sentido expuesto en la Sección 1.7, y su estimación se puede efectuar de manera más eficiente. Aunque algunos autores como Hirose y Lai (1997), los proponen como una estimación alternativa de un parámetro umbral, nunca los exploran desde el punto de vista de la verosimilitud perfil y mucho menos evalúan su robusticidad frente a cambios en los modelos considerados, como se verá en los ejemplos prácticos que se presentan en el Capítulo 4. Por otra parte, en la Sección 3.6 del Capítulo 3, se verá que las coberturas de los intervalos de verosimilitud perfil de un cuantil de probabilidad en el intervalo $(1/(n+1), n/(n+1))$ para una muestra de tamaño n , resultan ser muy buenas en la mayoría de los escenarios analizados.

Ahora, si lo que interesa es definitivamente la estimación de un umbral y se desean evitar los problemas de estimación que surgen con las verosimilitudes no acotadas, se pueden utilizar la verosimilitud restringida, presentada en la Sección 1.3.5 o bien la verosimilitud discretizada que se propone en el Capítulo 3. Ambas solucionan este tipo de problemas y son muy sencillas de implementar. El problema de verosimilitudes de forma aplanada en una región del espacio parametral, como la forma mostrada en la Figura 2.1.(c), es otro tipo de problema relacionado con la selección de modelos, como se verá en la Sección 2.4.

2.3 La verosimilitud restringida: una solución a verosimilitudes no acotadas

En esta sección se muestra que el problema de estimación ocasionado por verosimilitudes no acotadas que resultan de utilizar incorrectamente la aproximación continua a la verosimilitud, en densidades con singularidades, se resuelve al utilizar la verosimilitud restringida, presentada en la Definición 1.6. Para ilustrar esto considérese nuevamente el caso Weibull de mínimos. En el Ejemplo 2.2 se vio que la aproximación continua a la verosimilitud tomaba valores muy grandes al evaluarla en valores del umbral cercanos a la observación más pequeña. Si estas observaciones se

obtienen con un instrumento de medición que tiene una precisión $2h$, entonces al registrar la observación $Y_i = y_i$, en realidad se puede considerar que dicha observación cayó en el intervalo,

$$y_i - h < Y \leq y_i + h,$$

como se mencionó en la Sección 1.3.5.

Por tanto, es razonable considerar que al registrar la observación $y_{(1)}$, la variable aleatoria en realidad toma un valor en el intervalo $(y_{(1)} - h, y_{(1)} + h]$. En particular, es posible que la variable $Y_{(1)}$ tomase el valor $y_{(1)} - h$. Por tanto, habrá que considerar que los valores posibles que podría tomar el umbral μ son tales que $\mu \leq y_{(1)} - h$. Considerando lo anterior, la verosimilitud restringida se calcula de la siguiente manera,

$$\begin{aligned} L_R(\mu, \sigma, \beta; \mathbf{y}) &\propto \prod_{i=1}^n f(y_i; \mu, \sigma, \beta) I_{(-\infty, y_{(1)} - h]}(\mu) \\ &= \prod_{i=1}^n \frac{\beta}{\sigma} \left(\frac{y_i - \mu}{\sigma} \right)^{\beta-1} \exp \left[- \left(\frac{y_i - \mu}{\sigma} \right)^{\beta} \right] I_{(-\infty, y_{(1)} - h]}(\mu) \quad (2.3.1) \end{aligned}$$

donde $\sigma > 0$ y $\beta > 0$. Esta verosimilitud será siempre acotada, por construcción, puesto que evita evaluar la función de densidad en la singularidad que ocurre en $\mu = y_{(1)}$. La idea detrás de esta verosimilitud restringida es simple y elegante, además de que funciona muy bien para resolver algunos problemas de estimación que ocurren cuando $\beta < 1$ y se utiliza L_C . Las simulaciones presentadas en el Capítulo 3 corroboran esta afirmación y estudian el comportamiento de esta verosimilitud, bajo diferentes escenarios, comparándola a su vez con la verosimilitud discretizada que se propone en dicho capítulo. Por otra parte, ambas verosimilitudes se aplicarán a los ejemplos prácticos presentados en el Capítulo 4.

2.4 Verosimilitudes perfiles de forma aplanada

Existen ocasiones en que la forma de la verosimilitud perfil de un parámetro umbral tiene forma aplanada en cierta región del espacio parametral, como la mostrada en la Figura 2.1(c). En algunos casos esto se debe a que se cuenta con una muestra pequeña (Hirose y Lai, 1997), en otras se debe a un problema de identificabilidad de modelos (Montoya, 2008) y también ocurre cuando el modelo considerado está cerca de un modelo límite más simple. Algunos autores como Cheng y Iles (1990), también atribuyen esta situación a la forma particular de la distribución asimétrica de los datos, con cierto sesgo a uno de los lados. En cualquier caso, las observaciones no permiten distinguir o discriminar entre muchos de los miembros de la familia de modelos considerada.

En esta sección se describe el caso particular de familias de distribuciones que tienen un modelo límite más simple y en el Capítulo 4 se presentarán varios ejemplos de esta situación.

Las familias de distribuciones con un modelo límite más simple se dan para ciertos modelos de tres parámetros, de la forma $f(x - \mu, \theta)$, donde μ es un parámetro umbral y θ es el vector de parámetros restantes en el modelo. Con estos modelos y bajo ciertas condiciones, suelen obtenerse verosimilitudes perfiles para el parámetro umbral con una forma aplanada, en cierta región del espacio parametral. Cheng y Iles (1990) asocian esta situación con la existencia de modelos límite más simples, que describen bien a los datos y que resultan muy cercanos a los modelos de tres parámetros. A dichos modelos les llaman, en inglés, *embedded models* y distinguen que esta situación tiene un carácter diferente del caso de las verosimilitudes no-acotadas. Estos autores explican que para estos modelos de tres parámetros existe un modelo límite de dos parámetros cuyos parámetros son funciones de los parámetros originales, los cuales al estimarlos, toman valores muy grandes. En muchas ocasiones, por parsimonia, puede elegirse el modelo más simple para describir a los datos, con lo cual la estimación numérica de sus parámetros es mucho más eficiente; pero, como se muestra en los ejemplos analizados en el Capítulo 4, esto no debe considerarse una regla general.

En realidad, la situación de modelos límite es un caso particular del problema de selección de modelos estadísticos para un cierto conjunto de datos. Como se verá en la Sección 2.5.1 el modelo límite de dos parámetros para el modelo Weibull de tres parámetros resulta ser el modelo Gumbel. Éste también es el modelo límite para la familia de distribuciones Fréchet de tres parámetros.

Ahora, se ha visto que las distribuciones Gumbel, Weibull y Fréchet pueden escribirse como una sola distribución, conocida como Distribución Generalizada de Valores Extremos o DGVE. Esta distribución, para el caso de máximos, está dada por

$$G(y) = \begin{cases} \exp \left\{ - \left[1 + c \left(\frac{y-a}{b} \right)^{-\frac{1}{c}} \right] \right\}, & \text{si } c \neq 0, \\ \exp(-y) & \text{si } c = 0. \end{cases} \quad (2.4.1)$$

donde y es tal que $1 + c \left(\frac{y-a}{b} \right) > 0$, con $a \in \mathbb{R}$, $b > 0$ y $c \in \mathbb{R}$. En esta distribución a es un parámetro de localización, b es un parámetro de escala y c es un parámetro de forma. El caso $c < 0$ corresponde a la distribución Weibull y la relación de los parámetros de ésta con los de la DGVE es $\mu = a - \frac{b}{c}$, $\beta = -\frac{1}{c}$ y $\sigma = -\frac{b}{c}$. El caso $c > 0$ corresponde a la distribución Fréchet y la relación entre sus parámetros es $\mu = a - \frac{b}{c}$, $\beta = \frac{1}{c}$ y $\sigma = \frac{b}{c}$.

Para un modelo Weibull, cuando se tiene una verosimilitud relativa que tiene una forma aplanada alrededor del emv y es marcadamente asimétrica, se tendrá que la perfil del parámetro c bajo el modelo DGEV (con los mismos datos) asigna una verosimilitud alta o moderada al valor $c = 0$. Esto indica que un modelo Gumbel puede ser apropiado. Esto se explica fácilmente porque cuando c tiende a cero, la DGVE tiende a un modelo Gumbel. Lo mismo puede ocurrir con un modelo Fréchet el cual también tiene al modelo Gumbel como límite. Estas situaciones se analizan a detalle en los ejemplos incluidos en la Sección 4.3 del Capítulo 4.

El unificar estas tres familias en una sola familia simplifica mucho el aspecto de selección de modelos, ya que el parámetro de forma c proporciona información sobre el tipo de comportamiento de la cola de la distribución de máximos o mínimos. Esto permite, generalmente, seleccionar el tipo de distribución más apropiado para un conjunto de datos, sin necesidad de un juicio subjetivo. En Teoría de Extremos es muy común utilizar la DGVE ya que en muchas ocasiones se tiene gran incertidumbre acerca de la población madre de la que se extrajeron los máximos (o los mínimos).

Para la DGVE las propiedades asintóticas de los estimadores no se satisfacen automáticamente, pues las condiciones de regularidad que éstos requieren no se cumplen, ya que los extremos de esta distribución son funciones de los valores de los parámetros: $a - b/c$ es un extremo superior cuando $c < 0$ y un extremo inferior si $c > 0$. Así pues, las propiedades de los estimadores de máxima verosimilitud de la DGVE dependen del rango de valores que puede tomar el parámetro de forma c de esta distribución como exponen Coles (2004, p. 55) y Smith (1985). Existen críticas a la función de verosimilitud cuando $c < -1$, ya que se dice que los estimadores de máxima verosimilitud no existen. En realidad esto quiere decir que la aproximación continua a la verosimilitud no es válida ya que no está acotada. Cuando $c \in (-1, -0.5)$, se ha dicho que puede haber problemas de estimación, pues aunque se pueden obtener los estimadores de máxima verosimilitud, éstos no satisfacen las propiedades asintóticas estándares. Sin embargo, la mayoría de los datos que surgen en problemas prácticos, parecen cumplir la condición $-\frac{1}{2} \leq c \leq \frac{1}{2}$ (Kotz y Nadarajah, 2000, p.62).

2.5 Estimación en algunas densidades con parámetro umbral y una singularidad

En esta sección se presentan los problemas que surgen en la estimación de las densidades Weibull y lognormal de tres parámetros, que han dado lugar a críticas como las antes mencionadas y que son las distribuciones que se utilizan en los ejemplos prácticos que se incluyen en el Capítulo 4.

2.5.1 La distribución Weibull

Una familia de distribuciones de tres parámetros que es muy utilizada en un gran número de aplicaciones, por su flexibilidad al tomar una gran variedad de formas, es la distribución Weibull. Esta distribución toma el nombre del físico sueco Waloddi Weibull, quien la utilizó por vez primera en 1939 para modelar la fuerza de rompimiento de materiales; es él quien después lista una gran variedad de situaciones en las que puede utilizarse esta distribución. En la actualidad se considera muy adecuada para modelar tiempos de falla, la distribución de la fortaleza de ciertos materiales y en general para modelar una gran cantidad de comportamientos de tiempos de vida.

Al incluir un parámetro umbral en una distribución Weibull, éste puede representar una cota superior o inferior para la variable en estudio. Para el caso de mínimos, la función de densidad de una variable aleatoria Weibull de tres parámetros, es la dada por la Ecuación (2.1.3). Tal como se explicó en el Ejemplo 2.2, este modelo puede originar verosimilitudes no acotadas cuando $\mu \rightarrow y_{(1)}$ y con ello la aproximación continua a la verosimilitud puede presentar una forma de J , como la mostrada en la Figura 2.1(b). También es posible que con la distribución Weibull se originen verosimilitudes perfiles de forma aplanada como la presentada en la Figura 2.1(c).

En la Sección B.1 del Apéndice B, se muestra que el modelo Weibull de tres parámetros tiene a un modelo Gumbel como modelo límite de dos parámetros, lo cual se analiza estudiando el comportamiento de una densidad Weibull, cuando el parámetro de forma $\beta \rightarrow \infty$ o el parámetro umbral $\mu \rightarrow -\infty$. Para ello, se considera Y una variable aleatoria distribuida como una Weibull de tres parámetros, cuya función de densidad es la dada por (2.1.3). Para esta distribución Cheng y Iles (1990) sugieren la siguiente reparametrización uno a uno

$$(\mu, \sigma, \beta) \longleftrightarrow (\theta - \psi\beta, \psi\beta, \beta),$$

con la cual se puede demostrar que al tomar el límite cuando $\beta \rightarrow \infty$, se llega a

$$\lim_{\beta \rightarrow \infty} f(y; \theta, \psi, \beta) = \frac{1}{\psi} \exp\left(\frac{y - \theta}{\psi}\right) \exp\left[-\exp\left(\frac{y - \theta}{\psi}\right)\right],$$

para $-\infty < y < \infty$ que es la función de densidad de una variable aleatoria Gumbel.

Es muy común que se presenten problemas en la estimación de los parámetros de una distribución Weibull al usar métodos numéricos, por las consideraciones ya antes expuestas. Estos métodos son necesarios ya que no es posible obtener expresiones analíticas cerradas para todos los emv de los parámetros, como se muestra a continuación.

Consideremos la función de verosimilitud Weibull mostrada en la Ecuación (2.1.4). A partir de esta ecuación se calcula la log-verosimilitud siguiente:

$$l_C(\mu, \sigma, \beta; \mathbf{y}) = n \log \beta - n \log \sigma + (\beta - 1) \sum_{i=1}^n \log(y_i - \mu) - \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma}\right)^\beta,$$

de donde se obtienen las siguientes ecuaciones de verosimilitud,

$$\frac{\partial l_C(\mu, \sigma, \beta; \mathbf{y})}{\partial \mu} = -(\beta - 1) \sum_{i=1}^n (y_i - \mu)^{-1} + \sum_{i=1}^n \frac{\beta}{\sigma} \left(\frac{y_i - \mu}{\sigma}\right)^{\beta-1} = 0,$$

$$\frac{\partial l_C(\mu, \sigma, \beta; \mathbf{y})}{\partial \sigma} = -\frac{n\beta}{\sigma} + \frac{\beta}{\sigma} \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma}\right)^\beta = 0,$$

$$\frac{\partial l_C(\mu, \sigma, \beta; \mathbf{y})}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^n \log(y_i - \mu) - \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma}\right)^\beta \log\left(\frac{y_i - \mu}{\sigma}\right) = 0.$$

De ellas tan solo puede obtenerse una expresión para el parámetro de escala σ , que resulta

$$\hat{\sigma} = \left[\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^\beta \right]^{\frac{1}{\beta}}. \quad (2.5.1)$$

Nótese que que el cálculo de $\hat{\sigma}$ requiere el conocimiento de los otros dos parámetros, y como éstos no tienen una expresión cerrada, entonces, para la estimación de los tres parámetros son necesarios métodos de optimización numérica.

Los problemas de optimización están intrínsecamente asociados a los valores que toma el parámetro de forma β , sobre todo si este valor es menor a la unidad, porque entonces la verosimilitud perfil del umbral tendrá forma de J y será no acotada. En contraste, cuando el parámetro β es muy grande, la forma de la verosimilitud tiende a ser plana, como la presentada en la Figura 2.1(c). Esto presenta problemas de optimización debido a que la forma aplanada dificulta mucho dar con el máximo de la verosimilitud, de manera eficiente. Por ello va a ser de suma importancia arrancar estos métodos de búsqueda con buenos valores iniciales.

En la Figura 2.2 se pueden observar algunas formas que asume la densidad Weibull al variar el parámetro de forma β , en éstas se fijó el umbral en $\mu = 0$. Si $\beta = 3.5$ la densidad se aproxima bastante a una curva normal ya que el sesgo se vuelve cero, mientras que un valor $\beta = 0.5$ proporciona una curva sesgada a la derecha. Por otra parte, un valor de $\beta = 15$ origina una curva ligeramente sesgada a la izquierda. La distribución Weibull incluye a la distribución exponencial, cuando $\beta = 1$ y a la distribución Raleigh si $\beta = 2$. Un umbral diferente de cero simplemente traslada estas curvas a la derecha o izquierda, dependiendo si éste es positivo o negativo. Por otra parte, el parámetro de escala σ representa el percentil 63.2 de los datos y una simple transformación logarítmica, convierte una distribución Weibull en una distribución Gumbel.

A continuación se dan algunas sugerencias para la selección de los valores iniciales de los parámetros en el proceso de optimización numérica para la búsqueda de las estimaciones de máxima verosimilitud. Algunas de ellas las describen autores como Cheng y Amin (1983), Cheng y Iles (1990), Smith (1985), Mann (1984) y Heathcote et al. (2004), entre otros, y se pueden resumir en los puntos que enseguida se exponen, en los cuales también se describen situaciones que originan ciertos problemas de estimación.

1. Si $\beta > 2$ existirá un máximo local con la aproximación continua a la verosimilitud perfil del parámetro umbral. Este comportamiento puede asociarse a la forma de verosimilitud perfil mostrada en la Figura 2.1(a). En este caso la matriz de información de Fisher tiene todas las entradas finitas (Smith,

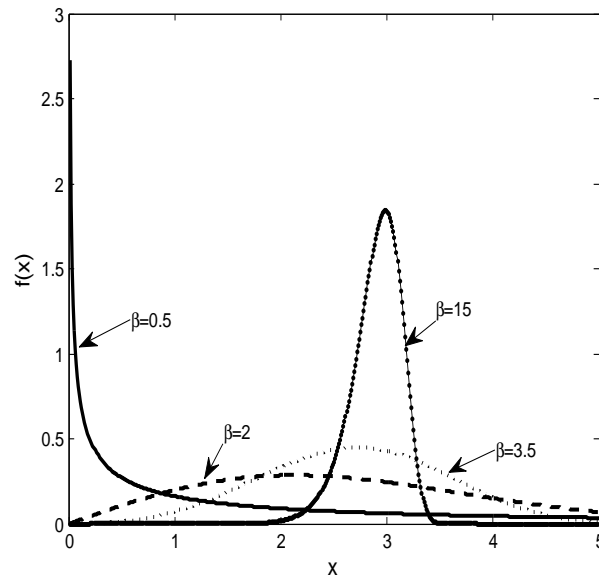


Figura 2.2: Formas de la densidad Weibull de mínimos para $\mu = 0$, $\sigma = 3$ y diferentes valores del parámetro de forma β .

1985) y se ha mostrado (Cheng y Iles, 1990) que son válidas las propiedades asintóticas de los estimadores de máxima verosimilitud.

2. Cuando $\beta \leq 2$ ciertos elementos de la matriz de información de Fisher toman valores infinitos y entonces ya no se cumplen las propiedades asintóticas de los estimadores de máxima verosimilitud. De acuerdo con Cheng y Amin (1983), si $1 < \beta \leq 2$ existirá un máximo local con la aproximación continua a la verosimilitud, que conduce a estimaciones consistentes para los parámetros, aspecto que también muestran Cheng y Iles (1990).
3. Cuando se tiene que $\beta < 1$, no existe máximo local. Smith (1985) muestra que no existe una sucesión consistente de estimadores del parámetro umbral y por ello, con los métodos iterativos generalmente se obtiene la estimación $\hat{\mu} = y_{(1)}$, lo que origina estimaciones inconsistentes para los otros parámetros.
4. El caso de $\beta = 1$ no presenta dificultad alguna, ya que resulta la distribución exponencial de dos parámetros, para la que se obtienen los estimadores de máxima verosimilitud $\hat{\mu} = y_{(1)}$ y $\hat{\sigma} = \bar{y} - y_{(1)}$.

Dado que para estimar los parámetros de una distribución Weibull son necesarios métodos numéricos de optimización, es importante utilizar valores iniciales adecuados. De igual manera se sabe que algunas reparametrizaciones de la función de verosimilitud pueden simetrizarla y con ello facilitar ampliamente el proceso de estimación de parámetros. A continuación se dan algunas sugerencias que se vio facilitan la estimación de los parámetros de un modelo Weibull con parámetro umbral.

- (a) Iniciar el proceso de estimación utilizando un modelo regular, como es un modelo Weibull de dos parámetros. Para ello puede restarse a la variable Y , que

se supone distribuida como una Weibull de tres parámetros, un conjunto de valores μ_0 que sean valores razonables para el parámetro umbral μ y luego estimar, por máxima verosimilitud, los parámetros σ y β del modelo, para cada uno de los valores de μ considerados. Esto facilita un proceso de optimización ya que una densidad Weibull de dos parámetros sólo tiene dos formas posibles. Si $\beta \leq 1$ es monótona decreciente y cuando $\beta > 1$ es unimodal y la moda se alcanza en $\sigma \left(\frac{\beta-1}{\beta}\right)^{1/\beta}$. Las estimaciones máximo verosímiles que se seleccionarán para los parámetros μ , σ y β , serán las que correspondan al valor de μ_0 con el cual se obtuvo el mayor valor de la verosimilitud.

- (b) Un modelo de localización y escala como el Gumbel puede también auxiliar en la estimación de parámetros de un modelo Weibull. Para ello es muy útil el conocer que si en un modelo Weibull el parámetro umbral $\mu = 0$, entonces la variable $Z = \log Y$ tiene una distribución Gumbel con parámetro de localización $\eta = \log \sigma$ y parámetro de escala $\psi = 1/\beta$. Ahora, si $\mu \neq 0$, entonces se puede utilizar $Z = \log(Y - \mu)$. Con ello los valores iniciales que se recomienda usar son $\sigma_0 = \exp(\eta)$ y $\beta_0 = 1/\psi$.
- (c) Cuando $\beta < 1$, la función de densidad Weibull tiene forma de J invertida. En estos casos es posible utilizar el dato mínimo $y_{(1)}$ como estimador del umbral y estimar los parámetros restantes suponiendo que se tiene un modelo Weibull de dos parámetros.
- (d) En ocasiones puede utilizarse un cuantil pequeño (o bien grande) como estimación alternativa a la del umbral. En el caso de mínimos y para la distribución Weibull de tres parámetros, la reparametrización a utilizar es la proporcionada en la ecuación (1.7.2), presentada en el Capítulo 1.

2.5.2 La distribución lognormal

Existen muchas situaciones donde las observaciones toman varios órdenes de magnitud y el histograma de ellas es sumamente asimétrico y con cola pesada a la derecha. Ejemplos de ello son la abundancia de especies, el período de latencia de una enfermedad infecciosa, la distribución de plomo en el aire, etcétera. La distribución lognormal suele describir bien estos fenómenos que se asocian con procesos independientes que se combinan multiplicativamente. La distribución de los datos es generalmente sesgada y tiene la característica de asumir solamente valores positivos. El incluir un parámetro umbral μ en una distribución lognormal, permite estimar una cota para los datos en estudio y la distribución simplemente se traslada μ unidades. Sin embargo, la simple inclusión de este parámetro puede complicar el proceso de estimación de sus parámetros, aunque nunca tanto como en el caso Weibull.

Una variable aleatoria Y se distribuye como una lognormal de tres parámetros, con parámetro umbral μ , si

$$\log(Y - \mu) \sim N(\gamma, \sigma^2), \quad (2.5.2)$$

donde γ es la media y σ es el parámetro de escala de la distribución normal correspondiente. La función de densidad de una variable que sigue una distribución lognormal de tres parámetros es

$$f(y; \mu, \gamma, \sigma) = \frac{1}{\sigma\sqrt{2\pi}(y-\mu)} \exp\left\{-\frac{1}{2\sigma^2} [\log(y-\mu) - \gamma]^2\right\} I_{(\mu, \infty)}(y), \quad (2.5.3)$$

donde $\sigma > 0$, $\mu \in \mathbb{R}$, y $\gamma \in \mathbb{R}$. Nótese que γ es también la mediana de la distribución lognormal. La función de distribución correspondiente, puede encontrarse numéricamente como

$$F(y; \mu, \gamma, \sigma) = \int_{\mu}^y f(t; \mu, \gamma, \sigma) dt. \quad (2.5.4)$$

En el caso lognormal de tres parámetros también es posible observar verosimilitudes perfiles no acotadas, tal como señalan Giesbrecht y Kempthorne (1976) y Hill (1963). Considérese que $\mathbf{y} = (y_1, \dots, y_n)$ son los valores que toma una muestra de variables aleatorias independientes, distribuidas como (2.5.3). La aproximación continua a la verosimilitud correspondiente es,

$$L_C(\mu, \gamma, \sigma; \mathbf{y}) \propto \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}(y_i - \mu)} \exp\left\{-\frac{1}{2\sigma^2} [\log(y_i - \mu) - \gamma]^2\right\}. \quad (2.5.5)$$

Cuando se fija un valor de μ , los estimadores restringidos de σ y γ son

$$\hat{\gamma}(\mu) = \frac{1}{n} \sum_{i=1}^n \log(y_i - \mu) \quad (2.5.6)$$

y

$$\hat{\sigma}^2(\mu) = \frac{1}{n} \sum_{i=1}^n [\log(y_i - \mu) - \hat{\gamma}(\mu)]^2. \quad (2.5.7)$$

Estos estimadores restringidos se pueden utilizar para obtener la verosimilitud perfil del parámetro umbral μ , la cual es

$$L_C^p(\mu; \mathbf{y}) = L(\mu, \hat{\gamma}(\mu), \hat{\sigma}(\mu); \mathbf{y}) \propto \prod_{i=1}^n \frac{1}{\hat{\sigma}(\mu)\sqrt{2\pi}(y_i - \mu)} \times \exp\left\{-\frac{1}{2\hat{\sigma}^2(\mu)} [\log(y_i - \mu) - \hat{\gamma}(\mu)]^2\right\}. \quad (2.5.8)$$

Claramente tiene una singularidad en $\mu = y_{(1)}$, heredada de la densidad misma, y motivo por el cual, cuando $\mu \rightarrow y_{(1)}$, la verosimilitud perfil tiende a infinito.

En el caso lognormal también es posible usar un cuantil adecuado como estimación alternativa a un parámetro umbral. Para ello se puede utilizar la reparametrización uno a uno,

$$(\mu, \gamma, \sigma) \longleftrightarrow (Q_{\alpha}, \gamma, \sigma),$$

donde $Q_\alpha = F^{-1}(\alpha; \mu, \gamma, \sigma)$. Como los cuantiles simetrizan, en gran medida, la función de verosimilitud, el proceso de optimización será más sencillo.

Al igual que en el caso Weibull, para la distribución lognormal y distribuciones como la gamma, log-logística y Gaussiana inversa, Cheng y Iles (1990) presentan las reparametrizaciones asociadas, que permiten encontrar el modelo límite para cada una de ellas. Para el caso lognormal de tres parámetros, este modelo límite es una distribución normal. Las distribuciones Gamma (de dos y de tres parámetros) y la Gaussiana inversa también tienen como modelo límite al normal.

Como se ha visto a través de este capítulo, las críticas que recibe la función de verosimilitud y que se relacionan con verosimilitudes no acotadas, pueden resolverse fácilmente utilizando la verosimilitud restringida o la verosimilitud exacta. Ésta última puede incorporar una precisión diferente en cada una de las mediciones. Sin embargo, el registro de la observación más pequeña, en el caso de mínimos, o la observación más grande, en el caso de máximos, requiere un tratamiento especial que hasta la fecha no se ha considerado y que lleva a utilizar una verosimilitud discretizada perfil, que se deriva de la definición de verosimilitud discretizada que se presenta en el siguiente capítulo y cuyo uso se sugiere para el caso de densidades que poseen un parámetro umbral y singularidades.

Capítulo 3

La función de verosimilitud discretizada

En este capítulo se atiende, en particular, el caso de densidades que poseen una singularidad y un parámetro umbral, lo cual en general dificulta el proceso de estimación de parámetros. Se mostrará la utilidad que en estos casos tiene la función de verosimilitud exacta, que se describió en la Sección 1.3.5, para variables aleatorias continuas y que incluye a la precisión del instrumento de medición. Se considerará que un instrumento de medición induce una partición en la recta real, que está directamente ligada a las mediciones que son observables con dicho instrumento. De esta manera, surge el concepto de muestra discretizada, que se definirá en este capítulo.

Al considerar modelos con un parámetro umbral que represente una cota inferior para la variable en estudio, la discretización de la observación más pequeña motiva la necesidad de incorporar una modificación a la función de verosimilitud exacta, que surge de concebir de una manera más adecuada, la probabilidad de observar el dato más pequeño. Así, cuando el parámetro umbral es una cota inferior para el soporte de la variable aleatoria, es posible que éste incluso pueda ser mayor que la observación registrada más pequeña. Por tanto, al calcular la verosimilitud perfil del umbral, debe tomarse en cuenta si el valor del umbral cae o no, dentro del intervalo que rodea a la observación más pequeña que fue registrada. Este aspecto nunca antes se ha considerado y se discutirá en detalle en la Sección 3.2. La verosimilitud discretizada que se define en dicha sección, toma en cuenta este aspecto, siempre está acotada y las estimaciones que se obtienen con ella están ligadas a la precisión del instrumento de medición.

En este capítulo se analizan diferentes aspectos de la verosimilitud discretizada, entre los cuales están las posibles formas que puede tomar la verosimilitud perfil discretizada de un parámetro umbral. Éstas se comparan con las formas que Smith y Weissman (1985) reportaron para la verosimilitud perfil de un parámetro umbral, cuando se utiliza la aproximación continua, mismas que fueron mostradas en la Sección 2.2.

En particular se analizarán dos casos no-regulares, que son el caso Weibull de tres parámetros, para mínimos y un caso particular de éste, que es la distribución exponencial con tiempo de vida garantizado. Aunque esta última distribución no posee singularidad alguna, se incluye debido a que es posible obtener expresiones analíticas cerradas para los estimadores de los parámetros, tanto con la aproximación continua

como con la verosimilitud discretizada. Sin embargo la distribución del estadístico de la razón de verosimilitud sólo es posible obtenerla, explícitamente, cuando se utiliza la aproximación continua a la verosimilitud, ya que aún en este caso tan simple, el cálculo de este estadístico se complica cuando se utilizan las expresiones analíticas de los estimadores obtenidos con la verosimilitud discretizada.

En cuanto a la distribución Weibull de tres parámetros, con ninguna de estas verosimilitudes es posible obtener expresiones analíticas para todos los estimadores de sus parámetros. Sin embargo, se puede establecer una similitud entre la verosimilitud discretizada y la distribución multinomial, que permite retomar los resultados asintóticos que han sido demostrados para esta última. En la Sección 3.6 se utilizan herramientas de simulación para estimar los parámetros de un modelo Weibull y se estudian las coberturas de los intervalos de verosimilitud para el caso de muestras moderadas y pequeñas.

3.1 El rol de la función de verosimilitud perfil en la inferencia científica

La función de verosimilitud perfil juega un papel de gran importancia en la inferencia científica, pues integra múltiples aspectos que son esenciales en la solución e interpretación de un problema. La verosimilitud perfil de un parámetro de interés permite evaluar, de manera sencilla, la robusticidad de las inferencias frente a cambios en los modelos considerados. Por ejemplo, al comparar las gráficas de la verosimilitud perfil de algún cuantil de interés, bajo los distintos modelos, es posible evaluar robusticidad, tal como se mostró en la Sección 1.6. En otras ocasiones puede corroborar la selección previa de una distribución o bien sugerir alguna otra. Tal es el caso de la verosimilitud perfil del parámetro de forma c de la DGVE o bien la perfil de un parámetro umbral, como se mostró en el Capítulo 2.

Por otra parte, con la verosimilitud perfil es sencillo utilizar reparametrizaciones uno a uno, que facilitan los procedimientos de estimación, como se explicó en la Sección 1.5. Gracias a ello se pueden usar reparametrizaciones específicas de interés o simplemente seleccionar alguna que simetrice la función de verosimilitud. En muchos casos es posible llegar a reparametrizaciones que resultan muy estables frente a cambios en los modelos, como es el caso de los cuantiles, expuesto en la Sección 1.7.

Es por ello que resulta de gran importancia el saber interpretar, dentro del contexto del problema que se esté analizando, toda la información que proporciona la verosimilitud perfil completa del parámetro en estudio.

La información que proporciona el error del instrumento de medición, así como la interpretación correcta del rango de valores que puede tomar un parámetro umbral, se aprovechan en su totalidad con la verosimilitud que se propone en este capítulo, y para la cual es necesario definir el concepto de muestra discretizada que se expone

a continuación.

3.2 Muestra discretizada

Muchos fenómenos naturales pueden ser descritos por medio de variables aleatorias que concebimos y denotamos como continuas. Sin embargo, cualquier instrumento de medición que se utilice para registrar las observaciones de estas variables aleatorias, siempre tendrá una precisión finita. Así, toda medición de una variable aleatoria continua está censurada a un intervalo, determinado por la precisión del instrumento de medición y por la observación misma registrada, que suele ser el punto medio de dicho intervalo. Esto fue identificado desde hace mucho tiempo por Barnard (1966) y Kempthorne (1966).

En la práctica no se cuenta con la medición genuina y de una variable aleatoria continua (que en realidad es latente) Y , sino con un valor x , que corresponde a la observación de una variable aleatoria discreta observable X , cuyos posibles valores son los puntos medios de los intervalos asociados al instrumento de medición utilizado. Las observaciones correspondientes a esta variable aleatoria discreta X son las que se usan para hacer inferencias acerca del modelo estadístico $f(y; \theta)$ en estudio.

A continuación se exponen dos maneras equivalentes de concebir este proceso de discretización. Ambas dan lugar a lo que denominaremos muestra discretizada, que es la muestra que se considera para todos los análisis efectuados en esta tesis.

Supóngase que Y es una variable aleatoria continua con distribución $F_Y(y; \theta)$, cuyo soporte W es un subconjunto de los números reales y considere que se utiliza un instrumento de medición con precisión $2h$, para registrar las observaciones de Y . Por facilidad considérese que la precisión de este instrumento es constante para todos los valores de Y , esto es, la precisión de éste no depende de la magnitud que tome Y . Cualquier instrumento con esta precisión induce una partición de la recta real, que consiste de una colección de intervalos disjuntos R_j , de la forma

$$R_j = (x_j - h, x_j + h] = ((2j + 1)h, (2j + 3)h], \quad (3.2.1)$$

donde j pertenece al conjunto de los números enteros \mathbb{Z} , y $x_j = (2j + 2)h$ es el punto medio del intervalo R_j . Sin pérdida de generalidad, se considerará que cada uno de estos intervalos es cerrado por la derecha y abierto por la izquierda. Es claro que los intervalos definidos por (3.2.1) satisfacen las condiciones siguientes:

a) $R_j \cap R_k = \emptyset$ para enteros $j \neq k$,

b) $\mathbb{R} = \bigcup_{j \in \mathbb{Z}} R_j$.

Cualquier instrumento de medición asociado a esta partición, registrará el valor x_j , siempre que la variable aleatoria continua Y caiga en el intervalo R_j . Esto es,

$$Y \in R_j \Leftrightarrow x_j - h < Y \leq x_j + h.$$

En ese caso, la variable discreta observable asociada X tomará el valor de x_j .

Por ejemplo, si se mide la altura de una persona con un instrumento de medición que tiene una precisión $2h = 1$ cm. y la altura registrada es de 165 cms., eso significa que la verdadera altura es un valor que cae dentro del intervalo R_{164} , donde

$$R_{164} = (164.5, 165.5] = (165 - h, 165 + h].$$

El valor 165, registrado, es el punto medio del intervalo R_{164} ; esto es $x_{164} = (2j + 2)h$ para $j = 164$. Ahora, si el instrumento tuviera una precisión $2h^* = 0.02$, entonces la misma observación tendría el índice $j = 8249$ y estaría asociada al intervalo

$$R_{8249} = (164.99, 165.01] = (165 - h^*, 165 + h^*],$$

perteneciente a una partición diferente en la recta real. Así, el índice j depende de la magnitud de la precisión $2h$ del instrumento de medición correspondiente. Nótese que siempre se cumple que $x_j = 0$ para $j = -1$ y cualquier valor positivo h . Cuando se tienen datos no-negativos se considera que el primer intervalo, en esta partición, está centrado en el cero.

Sea ahora \mathcal{F} , la familia de intervalos de la partición (3.2.1), esto es,

$$\mathcal{F} = \{R_j\}_{j \in \mathbb{Z}}.$$

Dado el soporte $W \subset \mathbb{R}$ de la variable aleatoria Y , existe al menos una subfamilia de \mathcal{F} que lo cubre, por lo que es posible hablar de la subfamilia más pequeña

$$\mathcal{F}' = \{R_{j_k}\}_{j_k \in A},$$

con esta propiedad, donde $A \subset \mathbb{Z}$. Esto es,

$$W \subset \bigcup_{j_k \in A} R_{j_k}.$$

Esta subfamilia \mathcal{F}' es la más pequeña en el sentido de que está contenida en cualquier otra subfamilia que contenga a W . Por ejemplo, en el caso de medir alturas de personas, éstas no pueden tomar valores negativos, lo que implica que el soporte W de Y es la parte no-negativa de la recta real, $W = \mathbb{R}^+ \cup \{0\}$. En este caso el conjunto A está dado por $A = \{-1, 0, 1, 2, \dots\}$.

Resumiendo, si Y_1, \dots, Y_n son variables aleatorias continuas, independientes e idénticamente distribuidas con la misma densidad $f_Y(y; \theta)$ y son medidas con un instrumento con precisión finita $2h$, entonces, las observaciones registradas se denotarán por x_1, x_2, \dots, x_n , donde x_i es el punto medio de un intervalo $R_{j_i} \in \mathcal{F}'$.

Entonces, para la variable aleatoria Y_i , $i = 1, \dots, n$, corresponde una medición que es el punto medio del intervalo correspondiente R_{j_i} , esto es, $x_i = (2j_i + 2)h$, para algún índice $j_i \in A$, o equivalentemente

$$Y_i \in R_{j_i} \Leftrightarrow [x_i - h < Y_i \leq x_i + h].$$

Este proceso de discretización puede concebirse de dos maneras diferentes:

1. La primera es considerar que las Y_i , $i = 1, \dots, n$, son variables aleatorias censuradas por intervalo, de las que solamente se sabe que el verdadero valor cae en el intervalo correspondiente R_{j_i} , para algún $j_i \in A$ y donde el valor registrado $x_i = (2j_i + 2)h$ es el punto medio de ese intervalo.
2. La segunda manera es considerar que cada variable aleatoria continua Y_i , $i = 1, \dots, n$, es latente y que tiene asociada una variable aleatoria discreta X_i , que sí es observable y que toma el valor x_i con probabilidad

$$\begin{aligned} P[X_i = x_i] &= P[x_i - h < Y_i \leq x_i + h] = \int_{x_i - h}^{x_i + h} f(y; \theta) dy \\ &= F_Y(x_i + h; \theta) - F_Y(x_i - h; \theta) = P[Y_i \in R_{j_i}]. \end{aligned}$$

En ambos casos, se tiene que $P[X_i = x_i] = P(Y_i \in R_{j_i})$ y ambas probabilidades dependen de la distribución de Y , y del parámetro desconocido θ . Nótese que en ambos casos se puede considerar un modelo multinomial, asociado a los intervalos con probabilidades positivas. En cualquiera de los esquemas anteriores, es posible que ocurran observaciones repetidas en la muestra observada x_1, \dots, x_n . En una muestra de tamaño n , el número esperado de observaciones repetidas que ocurren en el intervalo R_{j_i} es igual a $n \cdot P[Y_i \in R_{j_i}]$. El número de observaciones repetidas disminuye conforme el instrumento de medición utilizado es más preciso, esto es, cuando $2h$ es más pequeño.

Los dos escenarios antes expuestos dan lugar al concepto que se denominará *muestra discretizada* y que se define a continuación.

Definición 3.1 Muestra discretizada. Sean Y_1, \dots, Y_n variables aleatorias continuas, independientes e idénticamente distribuidas, con función de distribución $F_Y(y; \theta)$, donde $\theta \in \Theta \subset \mathbb{R}^k$. Supóngase que estas variables son observadas con un instrumento de medición que tiene una precisión $2h > 0$. Los valores (x_1, \dots, x_n) que el instrumento de medición registra para estas variables aleatorias continuas, se denomina muestra discretizada y corresponde al conjunto de puntos medios de intervalos de la forma

$$(x_i - h, x_i + h],$$

dentro de los cuales toma valores la variable aleatoria continua Y_i , para $i = 1, \dots, n$.

Todas las muestras que se trabajan en esta tesis se consideran resultado de mediciones que llevan implícitas la precisión de un instrumento de medición y son por lo

tanto muestras discretizadas.

En la siguiente sección se presenta una propuesta de verosimilitud, que utiliza el concepto de muestra discretizada, definido en esta sección y considera de una manera diferente, el conjunto de valores que puede asumir un parámetro umbral.

3.3 La verosimilitud discretizada

Los primeros en considerar que toda medición de una variable aleatoria continua Y está censurada dentro de un intervalo $(x - h, x + h]$, determinado por la medición discretizada x , registrada para esa variable y por la precisión $2h$ del instrumento de medición, fueron Barnard (1966) y Kempthorne (1966). Este último incorporó, explícitamente, la precisión del instrumento de medición en la función de verosimilitud de una variable aleatoria normal. Conceptualmente, es equivalente considerar que toda variable continua Y no es observable, sino latente, y que tiene una variable discreta asociada X , que sí es observable. Los valores que se registran para X dependerán del instrumento de medición y de su precisión, como se indicó en la Sección 3.2. Estos autores coinciden que cuando se considera lo anterior, en el cálculo de la función de verosimilitud, se evitan los problemas de estimación que surgen del uso de la aproximación continua a la verosimilitud, que resulta no-acotada en el caso de densidades que tienen singularidades.

Casi veinte años después, Kalbfleisch (1985) resaltó el hecho que, en el caso de variables aleatorias continuas debe tomarse en cuenta la precisión del instrumento de medición al calcular la función de verosimilitud, a la cual llamó verosimilitud exacta. También ejemplificó, por medio de la distribución exponencial, que la verosimilitud exacta y la aproximación continua son indistinguibles, siempre que el valor h sea pequeño en comparación a la magnitud del parámetro θ que se desea estimar. Esto es razonable dado que implica que se pide que el instrumento tenga una buena precisión para obtener observaciones que estimen bien a θ . Sin embargo, Kalbfleisch no comenta respecto a que la verosimilitud exacta resuelva problemas que se presentan con el uso de la aproximación continua a la verosimilitud, cuando las densidades tienen singularidades.

Por otra parte Meeker & Escobar (1998) y Lindsey (1999) concuerdan tanto con Barnard (1966) como con Giesbrecht y Kempthorne (1976), en que al utilizar la Definición 1.4, para calcular la verosimilitud de variables continuas, se resuelve el problema de verosimilitudes no-acotadas, que surge por el uso de la aproximación continua. Además, al definir la verosimilitud de esta manera se toma en cuenta la definición original que dio Fisher (1921), donde la verosimilitud es proporcional a la probabilidad de la muestra observada.

De esta manera, las inferencias acerca de un vector de parámetros desconocido θ en aquellos modelos que conocemos como continuos, forzosamente se basan en una

muestra discretizada como la definida anteriormente. En ésta, los valores registrados x_1, \dots, x_n , pueden concebirse como mediciones de variables aleatorias discretizadas X_1, \dots, X_n , o bien como una realización de variables aleatorias continuas Y_1, \dots, Y_n , censuradas por intervalo. En cualquiera de los dos casos la función de verosimilitud asociada $L(\theta; x_1, \dots, x_n)$ se puede expresar de la manera siguiente

$$\begin{aligned} L(\theta; x_1, \dots, x_n) &= C(x_1, \dots, x_n) \prod_{i=1}^n P[X_i = x_i; \theta] \\ &= C(x_1, \dots, x_n) \prod_{i=1}^n P[x_i - h < Y_i \leq x_i + h; \theta] \\ &= C(x_1, \dots, x_n) \prod_{i=1}^n [F_Y(x_i + h; \theta) - F_Y(x_i - h; \theta)], \quad (3.3.1) \end{aligned}$$

donde $C(x_1, \dots, x_n)$ es una función no-negativa y acotada, que depende solamente de las observaciones registradas y no de los parámetros desconocidos. Nótese que esta verosimilitud siempre será acotada. Por tanto, evita el problema que presenta la aproximación continua a la verosimilitud, la cual no es acotada cuando se utiliza en densidades que cuentan con un parámetro umbral y una singularidad.

Sin embargo, no es claro cómo calcular la probabilidad asociada al intervalo que corresponde a la observación más pequeña, cuando el parámetro umbral es una cota inferior; o bien el asociado a la observación más grande, cuando el umbral es una cota superior. Es importante aclarar esta situación pues teóricamente el parámetro umbral, que denotaremos por μ , delimita el soporte de la variable aleatoria continua Y .

Cuando el umbral es una cota inferior, debe cumplirse que $\mu \leq Y$. Conviene entonces, aclarar cuál es la relación de orden que guardan el valor más pequeño registrable $x_{(1)}$ y el umbral μ . Esto es, la variable aleatoria continua Y puede tomar en realidad cualquier valor en el intervalo que rodea a la observación discretizada más pequeña, $(x_{(1)} - h, x_{(1)} + h]$. Por tanto es posible que el umbral μ sea inclusive mayor que esta observación $x_{(1)}$. En ese caso se tendría que

$$x_{(1)} < \mu \leq Y \leq (x_{(1)} + h).$$

Esta situación nunca se ha considerado con anterioridad en la literatura estadística, por lo que a continuación se explica con más detalle.

Sea $\theta = (\varphi, \mu) \in \Theta \subset \mathbb{R}^k$ un vector de parámetros, donde μ es un parámetro umbral y φ es el vector de parámetros restante. Supóngase ahora que se tienen tres variables aleatorias continuas Y_1, Y_2, Y_3 que son independientes e idénticamente distribuidas como $F(y; \theta)$. Las mediciones correspondientes, que conforman la muestra discretizada observada son x_1, x_2 y x_3 . Ahora, las observaciones ordenadas son: $x_{(1)} \leq x_{(2)} \leq x_{(3)}$, donde $x_{(i)}$ no necesariamente es igual a x_i , como puede observarse en la Figura 3.1.

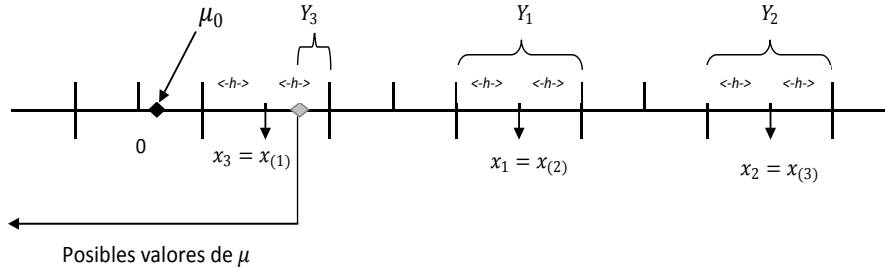


Figura 3.1: Muestra discretizada y su variable aleatoria continua asociada. Se señalan los posibles valores que puede tomar μ , tanto dentro del intervalo que rodea $x_{(1)}$, como a la izquierda de éste. Se muestra también el verdadero valor μ_0 , desconocido, del umbral.

En esta figura se puede observar que es posible que el valor que toma la variable aleatoria Y_3 , que origina la medición discretizada $x_3 = x_{(1)}$, sea tal que $Y_3 \geq x_{(1)}$, por lo que el parámetro umbral puede tomar valores que satisfagan,

$$x_{(1)} \leq \mu \leq Y_3 \leq x_{(1)} + h,$$

y ser más grande que la observación más pequeña que se registró. En la misma figura puede observarse que el parámetro umbral μ pudiera también tomar valores a la izquierda de $x_{(1)} - h$. Es por ello que, como se dijo anteriormente, la contribución que a la verosimilitud hace el intervalo asociado a la observación más pequeña, depende de considerar si el parámetro umbral μ se encuentra o no dentro del intervalo $(x_{(1)} - h, x_{(1)} + h]$. Esta contribución puede resumirse como,

$$P [x_{(1)} - h < Y \leq x_{(1)} + h] = \begin{cases} \int_{\mu}^{x_{(1)}+h} f_Y(t; \varphi, \mu) dt, & \text{si } \mu \in (x_{(1)} - h, x_{(1)} + h] \\ \int_{x_{(1)}-h}^{x_{(1)}+h} f_Y(t; \varphi, \mu) dt, & \text{si } \mu \leq x_{(1)} - h \\ 0, & \text{si } \mu > x_{(1)} + h, \end{cases}$$

donde f_Y es la función de densidad asociada a las variables aleatorias Y_1, \dots, Y_n .

Así pues, cuando se hacen inferencias acerca de los parámetros de un modelo cuya densidad tiene un parámetro umbral, no sólo debe considerarse la precisión del instrumento de medición, sino también la posibilidad de que el parámetro umbral pueda caer dentro o fuera del intervalo que rodea a la observación más pequeña.

Esta posibilidad conduce a proponer una modificación de la verosimilitud presentada en (3.3.1). En esta nueva definición se utiliza una muestra discretizada ya ordenada $(x_{(1)}, \dots, x_{(n)})$, y se calcula, de manera especial, la contribución que el intervalo $(x_{(1)} - h, x_{(1)} + h]$ aporta a la verosimilitud. Debe recordarse que tanto en la Definición 3.1, de muestra discretizada, como en la definición de verosimilitud discretizada que se presenta enseguida, por sencillez se considera que la precisión del instrumento de medición es siempre constante e igual a $2h$. Aún cuando cada observación tuviera una precisión diferente, es posible utilizar la función de verosimilitud discretizada propuesta. Tal es el caso del ejemplo presentado en la Sección

4.1.3 relativo al índice de masa corporal de mujeres sonorenses, donde cada observación tiene asociado un error de medición particular, que se calcula a través de un procedimiento muy utilizado en física y que se conoce como propagación de errores.

Definición 3.2 *Verosimilitud discretizada para variables aleatorias continuas cuya densidad tiene un parámetro umbral y una singularidad.* Sean Y_1, Y_2, \dots, Y_n variables aleatorias continuas, independientes e idénticamente distribuidas con función de densidad $f(y; \theta)$ y función de distribución $F(y; \theta)$, donde $\theta = (\varphi, \mu) \in \Theta \subset \mathbb{R}^k$, μ un parámetro umbral, tal que $Y_i \geq \mu$ para $i = 1, \dots, n$; y φ el vector de parámetros restantes. Sea $x_{(1)}, \dots, x_{(n)}$ la muestra discretizada ordenada, asociada a estas variables aleatorias y obtenida con un instrumento de medición con precisión $2h$. La verosimilitud discretizada para θ , $L_D : \Theta \rightarrow \mathbb{R}$, es una función que es proporcional a la probabilidad de la muestra observada y se define como

$$L_D(\varphi, \mu; x_{(1)}, \dots, x_{(n)}) = C(x_1, \dots, x_n) \prod_{i=1}^n P[x_{(i)} - h < Y \leq x_{(i)} + h; \varphi, \mu]$$

donde $C(x_1, \dots, x_n)$ es una función acotada, no-negativa, que depende solamente de la muestra y

$$P[x_{(1)} - h < Y \leq x_{(1)} + h] = \begin{cases} \int_{\mu}^{x_{(1)}+h} f_Y(t; \varphi, \mu) dt, & \text{si } \mu \in (x_{(1)} - h, x_{(1)} + h) \\ \int_{x_{(1)}-h}^{x_{(1)}+h} f_Y(t; \varphi, \mu) dt, & \text{si } \mu \leq x_{(1)} - h \\ 0, & \text{si } \mu > x_{(1)} + h. \end{cases}$$

La verosimilitud discretizada L_D tiene la propiedad de que siempre está acotada, ya que está definida como un producto de probabilidades, a diferencia de la aproximación continua a la verosimilitud que hereda las singularidades que tenga la función de densidad. Ahora, como L_D está acotada por arriba, el supremo existe y por tanto siempre existe el emv, si bien no necesariamente es único.

A continuación se presenta cómo calcular la verosimilitud discretizada para el caso Weibull de tres parámetros.

Ejemplo 3.1 Sea $\mathbf{x} = (x_1, \dots, x_n)$, una muestra discretizada que resulta de medir n variables aleatorias continuas independientes, Y_1, \dots, Y_n que se distribuyen como una Weibull de mínimos, con parámetro umbral μ , parámetro de escala $\sigma > 0$, parámetro de forma $\beta > 0$ y cuya función de distribución es

$$F(y; \mu, \sigma, \beta) = \left\{ 1 - \exp \left[- \left(\frac{y - \mu}{\sigma} \right)^\beta \right] \right\} I_{[\mu, \infty)}(y).$$

Para calcular la verosimilitud discretizada L_D de la Definición 3.2, es necesario considerar la posibilidad de que el parámetro umbral pueda tomar valores tanto dentro como fuera del intervalo que rodea a la observación discretizada más pequeña $x_{(1)}$. Es por ello que L_D se expresa como:

$$L_D(\mu, \sigma, \beta; \mathbf{x}) = \begin{cases} L_1(\mu, \sigma, \beta; \mathbf{x}), & \text{si } \mu \leq x_{(1)} - h \\ L_2(\mu, \sigma, \beta; \mathbf{x}), & \text{si } \mu \in (x_{(1)} - h, x_{(1)} + h) \\ 0, & \text{si } \mu > x_{(1)} + h. \end{cases} \quad (3.3.2)$$

Así la función L_1 definida como

$$L_1(\mu, \sigma, \beta; \mathbf{x}) = \prod_{i=1}^n [F(x_{(i)} + h) - F(x_{(i)} - h)],$$

considera la posibilidad de que el parámetro umbral μ tome solamente valores fuera del intervalo que rodea a $x_{(1)}$. A su vez, la función L_2 definida como

$$L_2(\mu, \sigma, \beta; \mathbf{x}) = [F(x_{(1)} + h) - F(\mu)]^m \prod_{i=m+1}^n [F(x_{(i)} + h) - F(x_{(i)} - h)]$$

toma en cuenta la posibilidad que el umbral pueda tomar valores en $(x_{(1)} - h, x_{(1)} + h]$, denotando por m el número de veces que se repite la observación más pequeña. Al igual que en el caso de utilizar la aproximación continua a la función de verosimilitud, no es posible obtener, analíticamente, todos los estimadores de los parámetros del modelo Weibull, por lo que son necesarios métodos numéricos de optimización.

Puede observarse que los componentes de la Ecuación (3.3.2) son probabilidades y por tanto siempre serán acotadas, por lo que no hay problemas para la estimación de parámetros, por máxima verosimilitud. Sin embargo, el caso de verosimilitudes perfiles con formas aplanadas puede aún presentarse ya que se trata de un problema de selección de modelos, como se explicó en la Sección 2.4.

El cálculo de la verosimilitud discretizada en el caso Lognormal de tres parámetros se realiza de una manera similar al caso Weibull, con la diferencia de que la función de distribución es la dada por la ecuación (2.5.4), presentada en la Sección 2.5.2.

Para calcular una verosimilitud perfil, a partir de la verosimilitud discretizada, se procede de manera similar a lo presentado en la Sección 1.3.8, como se muestra en la siguiente definición.

Definición 3.3 Verosimilitud discretizada perfil. Sea $L_D(\theta; \mathbf{y})$ la función de verosimilitud discretizada para $\theta = (\lambda, \psi)$. La verosimilitud discretizada perfil de λ , que se denotará por $L_D^p(\lambda; \mathbf{y})$, se define como

$$L_D^p(\lambda; \mathbf{y}) = \sup_{\psi|\lambda} L_D(\lambda, \psi; \mathbf{y}) = L_D(\lambda, \hat{\psi}(\lambda; \mathbf{y}); \mathbf{y}), \quad (3.3.3)$$

donde $\hat{\psi}(\lambda; \mathbf{y})$ es el estimador de máxima verosimilitud restringido de ψ , que se obtiene maximizando la función de verosimilitud discretizada para un valor fijo de λ .

En esta tesis, la verosimilitud relativa perfil que se calcula a partir de la función de verosimilitud discretizada, se denotará por R_D^p y para el caso de la verosimilitud restringida perfil, utilizaremos la notación R_r^p .

3.3.1 Relación de la verosimilitud discretizada con otras verosimilitudes

En esta sección se analizan las similitudes y diferencias que existen entre la verosimilitud discretizada, propuesta en la sección anterior y otras verosimilitudes como son la verosimilitud para datos censurados por intervalo y la verosimilitud para datos agrupados, presentadas en las Secciones 1.3.6 y 1.3.7, así como la verosimilitud exacta de la Definición 1.4. La diferencia básica entre la verosimilitud discretizada y estas otras verosimilitudes radica en que la verosimilitud discretizada se concibe para densidades que poseen un parámetro umbral y una singularidad, y la probabilidad asociada al intervalo correspondiente a la observación discretizada más pequeña (o bien, la más grande), se calcula de manera diferente, como se expuso en la sección anterior.

La verosimilitud para datos censurados por intervalo es comparable con la verosimilitud exacta, ya que, matemáticamente, la probabilidad que cada uno de los intervalos aporta a la función de verosimilitud, se calcula de la misma manera. Si la amplitud de cada uno de estos intervalos coincide con la precisión del instrumento de medición utilizado, entonces serán equivalentes. El problema de datos censurados se ha investigado ampliamente, tanto desde el punto de vista paramétrico como no-paramétrico. En general, la forma de analizarlo depende de la información con la que se cuente. Para ambos enfoques se han presentado diversos resultados en la literatura estadística. Autores como Odell et al. (1992) proponen, por ejemplo, utilizar los puntos medios de los intervalos de censura como si fuesen los datos observados y utilizar las técnicas comunes de máxima verosimilitud, en la estimación de parámetros. Sin embargo, Gómez et al. (2004) argumentan que este procedimiento puede generar sesgo en las estimaciones. En realidad, existen muchos estudios relativos a muestras censuradas y algunos de los resultados obtenidos podrían extrapolarse, en ciertos casos, a la verosimilitud discretizada, siempre y cuando se satisfagan los supuestos que se establecen para su validez.

En cuanto a la verosimilitud agrupada, Giesbrecht y Kempthorne (1976) la definen de manera equivalente a lo que Kalbfleisch (1985) llamó verosimilitud exacta. Ellos analizan el caso lognormal de tres parámetros y al no contar con estimadores explícitos para todos los parámetros de esta distribución, estiman las probabilidades asociadas a cada uno de los intervalos de agrupamiento (cuya amplitud es igual a la precisión del instrumento de medición) y sus estimadores los consideran comparables a los que se obtienen, por máxima verosimilitud, en el caso discreto multinomial. Ellos muestran que las propiedades asintóticas que se cumplen para el caso multinomial, son aplicables también a la verosimilitud agrupada. De manera equivalente, se pueden retomar estos resultados para el caso de la verosimilitud discretizada. Lo anterior se sustenta en lo establecido por el Principio de Verosimilitud (Cox y Hinkley, 2000, p. 39), ya que si dos funciones de verosimilitud son proporcionales, forzosamente deben conducir a las mismas conclusiones.

Aunque los resultados asintóticos demostrados para otras verosimilitudes, se

puedan utilizar con la verosimilitud discretizada, es importante señalar que muchos problemas reales involucran, por lo general, muestras pequeñas. Por ello la importancia de conocer el comportamiento de la verosimilitud propuesta, cuando no se cuenta con muestras grandes. Como se muestra en la Sección 3.6, la metodología estadística que se propone en esta tesis, permite dar intervalos de verosimilitud-confianza con coberturas buenas, incluso en el caso de muestras moderadas o pequeñas. Solamente cuando se estima un parámetro umbral, se requieren muestras de mayor tamaño para lograr buenas coberturas. Esto ocurre, particularmente, cuando la densidad tiene forma convexa y la probabilidad de que ocurran observaciones cercanas al umbral aumenta.

El proceso de simulación efectuado en esta tesis resultó muy relevante e informativo, ya que permitió caracterizar las propiedades de cobertura que, bajo diferentes escenarios, poseen los intervalos de verosimilitud discretizada de los parámetros de un modelo estadístico Weibull, cuando se tienen muestras pequeñas y moderadas. Se seleccionó este modelo ya que es ampliamente utilizado y por otra parte, es muy común que al utilizarlo se cuente con muestras pequeñas, problema que no había sido analizado bajo este enfoque. Los resultados que se presentan solamente pueden obtenerse por medio de la estadística computacional, que hoy en día juega un papel fundamental en el desarrollo de la estadística, como se explica en la siguiente sección.

3.4 La relevancia de la estadística computacional en las inferencias para muestras pequeñas

Existen diferencias fundamentales entre la solución analítica y la solución numérica de un problema. La solución analítica proporciona una prueba matemática precisa y como tal, no siempre puede encontrarse. En cambio, una solución numérica generalmente puede hallarse pero muchas veces lleva implícita un cierto grado de error. Por ejemplo, bajo la mayoría de los enfoques de estimación no es posible obtener expresiones analíticas cerradas para todos los estimadores puntuales de los parámetros de un modelo Weibull. Por tanto, se dificulta estudiar analíticamente las propiedades o el comportamiento asintótico de dichos estimadores. Sin embargo, siempre es posible realizar inferencias acerca de los parámetros de este tipo de modelos utilizando métodos numéricos computacionales que forman parte de lo que se conoce como *estadística computacional*.

La estadística computacional se refiere al diseño de algoritmos que permiten implementar, por medio del uso de computadoras, métodos estadísticos que serían imposibles de realizar de otra manera. Ejemplo de ello son los métodos de remuestreo (Bootstrap), simulaciones diversas, métodos de cadenas de Markov y métodos Monte Carlo (MCMC), entre otros. Estas técnicas permiten resolver, computacionalmente, muchos problemas que resultan analíticamente intratables (Lauro, 1996). En la actualidad existen diversas áreas de investigación dentro de la estadística computacional, como son la minería de datos, la construcción y evaluación de modelos, técnicas

para descubrir estructuras en los datos, tanto de manera exploratoria como visual, técnicas de remuestreo, métodos de simulación, modelación estadística, etcétera.

En esta tesis se han utilizado, en gran medida, simulaciones y métodos de optimización numérica para poder implementar y valorar la propuesta de la verosimilitud discretizada, que se define en la Sección 3.3, en modelos estadísticos que tienen un parámetro umbral y una singularidad. Para verificar que esta propuesta es razonable, práctica y sencilla, para estimar parámetros en estos casos no-regulares, se creó un algoritmo computacional que implementa lo descrito en la Definición 3.2. Los resultados obtenidos se analizaron desde diferentes aspectos. Por ejemplo, para verificar qué tan buenas pueden ser las estimaciones obtenidas con la verosimilitud discretizada se calculó la frecuencia de cobertura de ciertos intervalos de verosimilitud. El analizar las condiciones bajo las que ocurren coberturas bajas, permitió identificar los tipos de modelos en los que resulta difícil estimar los parámetros, cuando se tienen muestras pequeñas.

Así pues, los resultados que se presentan en esta tesis son producto, en su mayoría, del intenso uso de herramientas que forman parte de la estadística computacional, mismas que hoy en día son ampliamente utilizadas para validar procedimientos y proporcionar orientación en problemas tanto prácticos como teóricos (Gentleman, 2004).

Tan sólo con el fin de mostrar el tipo de dificultades teóricas que surgen al analizar un modelo no-regular, se muestra el caso particular de determinar el comportamiento del estadístico de la razón de verosimilitudes, para un caso particular del modelo Weibull. Este estadístico es de gran utilidad para asignar probabilidades de cobertura a los intervalos de verosimilitud perfil. Sin embargo, calcular analíticamente su distribución puede resultar muy complicado, como se muestra en la siguiente sección donde, con fines comparativos, primero se utiliza la aproximación continua a la verosimilitud y después se usa la verosimilitud discretizada propuesta.

3.5 Distribución del estadístico de la razón de verosimilitudes para un caso no-regular

Conocer la distribución exacta del estadístico de razón de verosimilitud, aún en casos regulares, es una situación afortunada ya que este resultado depende de aproximaciones a muestras grandes. Cuando un modelo es no-regular, la distribución asintótica del estadístico de la razón de verosimilitud por lo general sufre modificaciones y su distribución puede ser diferente a la obtenida en los casos regulares.

La distribución exponencial con tiempo de vida garantizado incluye un parámetro umbral que delimita el soporte de la variable aleatoria, por lo que se clasifica como un modelo no-regular. Es un caso particular de la distribución Weibull, que se utiliza a través de toda la tesis, y fue por tanto, el modelo seleccionado para mostrar

las dificultades analíticas que pueden presentarse al analizar modelos no-regulares.

La función de densidad de una variable aleatoria Y con distribución exponencial con tiempo de vida garantizado, parámetro umbral μ y parámetro de escala $\sigma > 0$ es

$$f(y; \mu, \sigma) = \frac{1}{\sigma} \exp\left[-\frac{(y - \mu)}{\sigma}\right] I_{[\mu, \infty)}(y),$$

y su función de distribución está dada por

$$F(y; \mu, \sigma) = 1 - \exp\left[-\frac{(y - \mu)}{\sigma}\right] I_{[\mu, \infty)}(y). \quad (3.5.1)$$

En esta distribución, el parámetro μ puede interpretarse como un valor a partir del cual inician los tiempos de vida, o bien un valor que garantiza que no ocurrirán fallas antes de ese punto. Este parámetro traslada el inicio de la distribución, una distancia μ a la derecha del origen. Así, la distribución inicia en $y = \mu$, es convexa y decrece monótonamente conforme y se aleja de μ .

Con el fin de tener un punto de comparación de los resultados obtenidos para la distribución del estadístico de la razón de verosimilitud, cuando se utiliza la verosimilitud discretizada, se calcula primeramente los resultados que se obtienen al utilizar la aproximación continua a la verosimilitud.

3.5.1 Aproximación continua a la verosimilitud para la exponencial con tiempo de vida garantizado (t.v.g.)

Para obtener analíticamente, la distribución asintótica del estadístico de la razón de verosimilitud, es necesario contar con una expresión analítica para los estimadores de los parámetros. En este caso necesitamos conocer los estimadores que se obtienen para el parámetro umbral μ y el parámetro de escala $\sigma > 0$, cuando se utiliza la aproximación continua a la función de verosimilitud. Para ello consideremos que $\mathbf{y} = (y_1, y_2, \dots, y_n)$ es una muestra aleatoria observada, de tamaño n , de una distribución exponencial con parámetros μ y σ ; la aproximación continua a la verosimilitud se calcula como

$$L(\mu, \sigma; \mathbf{y}) \propto \prod_{i=1}^n \frac{1}{\sigma} \exp\left[-\frac{(y_i - \mu)}{\sigma}\right] I_{(-\infty, y_{(1)}]}(\mu) I_{(0, \infty)}(\sigma),$$

de donde se obtiene que los emv de μ y de σ son $\hat{\mu} = y_{(1)}$ y $\hat{\sigma} = \bar{y} - \hat{\mu}$, respectivamente.

Para estudiar el comportamiento de $-2 \log R(\mu_0; \mathbf{y})$, donde μ_0 se considerará el valor real de μ , supondremos, para facilitar los cálculos, que σ es conocida, por lo

que se sustituirá por $\sigma = \sigma_0$.

$$\begin{aligned} R(\mu_0; \mathbf{y}) &= \frac{L(\mu_0, \sigma_0; \mathbf{y})}{L(\hat{\mu}, \sigma_0; \mathbf{y})} \\ &= \prod_{i=1}^n \exp\left[-\frac{(y_i - \mu_0)}{\sigma_0}\right] \left\{ \prod_{i=1}^n \exp\left[-\frac{(y_i - \hat{\mu})}{\sigma_0}\right] \right\}^{-1} \\ &= \exp\left[\frac{n(\mu_0 - y_{(1)})}{\sigma_0}\right] \end{aligned} \quad (3.5.2)$$

Así, tendremos que

$$\begin{aligned} P(-2 \log R(\mu_0; \mathbf{y}) \leq z) &= P\left\{-2 \log \cdot \exp\left[\frac{n(\mu_0 - y_{(1)})}{\sigma_0}\right] \leq z\right\} \\ &= P\left(y_{(1)} \leq \mu_0 + \frac{z\sigma_0}{2n}\right) \\ &= 1 - \exp\left(-\frac{z}{2}\right). \end{aligned}$$

Ahora, como la función de densidad de una variable aleatoria Z que se distribuye χ_k^2 es

$$f(z) = \frac{z^{\frac{k}{2}-1} \exp\left\{-\frac{z}{2}\right\}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}},$$

se deduce que $-2 \log R(\mu_0; \mathbf{y})$ sigue una distribución $\chi_{(2)}^2$, pues $\Gamma(1) = 1$.

Un estudio detallado del comportamiento del estadístico de la razón de verosimilitudes, para esta distribución, lo muestra Lawless (2003, p. 190).

3.5.2 Verosimilitud discretizada para la exponencial con t.v.g.

En esta sección se presentan algunos resultados que se obtienen al calcular, de manera analítica, la distribución del estadístico de la razón de verosimilitudes, cuando se utiliza la verosimilitud discretizada L_D , en la distribución exponencial con tiempo de vida garantizado. Como se podrá ver, el llegar a una expresión para este estadístico, siguiendo el procedimiento mostrado en la Sección 3.5.1, no es tan directo y depende de ciertas condiciones.

Para el cálculo de la función de verosimilitud discretizada se requiere usar la función de distribución de la exponencial con t.v.g., presentada en (3.5.1). Para ello considérese que $\mathbf{x} = (x_{(1)}, \dots, x_{(n)})$ es una muestra discretizada ordenada, que se obtuvo con un instrumento de medición que posee una precisión fija $2h$. Luego, la verosimilitud discretizada $L_D(\mu, \sigma; \mathbf{x})$, basada en esta muestra discretizada es

$$L_D(\mu, \sigma; \mathbf{x}) = L_1(\mu, \sigma; \mathbf{x}) + L_2(\mu, \sigma; \mathbf{x}) \quad (3.5.3)$$

donde

$$L_1(\mu, \sigma; \mathbf{x}) = \exp\left(-\sum_{i=1}^n \frac{x(i)}{\sigma}\right) \times \left[\exp\left(\frac{\mu+h}{\sigma}\right) - \exp\left(\frac{\mu-h}{\sigma}\right)\right]^n I_{(-\infty, x_{(1)}-h]}(\mu) I_{(0, \infty)}(\sigma) \quad (3.5.4)$$

y

$$L_2(\mu, \sigma; \mathbf{x}) = \left\{1 - \exp\left[-\frac{(x_{(1)}+h-\mu)}{\sigma}\right]\right\}^m \exp\left(-\sum_{i=m+1}^n \frac{x(i)}{\sigma}\right) \times \left[\exp\left(\frac{\mu+h}{\sigma}\right) - \exp\left(\frac{\mu-h}{\sigma}\right)\right]^{n-m} I_{(x_{(1)}-h, x_{(1)}+h]}(\mu) I_{(0, \infty)}(\sigma). \quad (3.5.5)$$

La verosimilitud (3.5.3) debe expresarse como una suma de las funciones (3.5.4) y (3.5.5), cuyos soportes son excluyentes y dependientes del parámetro μ , ya que este umbral puede vivir dentro o fuera del intervalo $x_{(1)} \pm h$, lo cual dependerá de la muestra que se esté analizando.

Ahora, tanto μ_0 , el valor verdadero del parámetro μ , como $\hat{\mu}$, su estimador máximo verosímil, pueden estar dentro o fuera del intervalo que rodea a la observación más pequeña $x_{(1)}$. Considerando lo anterior, se tienen cuatro escenarios posibles que se presentan en la siguiente tabla.

	$\mu_0 \in (-\infty, x_{(1)} - h]$	$\mu_0 \in (x_{(1)} - h, x_{(1)} + h]$
$\hat{\mu} \in (-\infty, x_{(1)} - h]$	Caso 1	Caso 2
$\hat{\mu} \in (x_{(1)} - h, x_{(1)} + h]$	Caso 3	Caso 4

Tabla 3.1: Escenarios posibles para el cálculo de L_D en el modelo exponencial con t.v.g.

Las expresiones analíticas de los estimadores de los parámetros μ y σ dependen, tanto del escenario que se esté considerando, como de la precisión del instrumento de medición. De acuerdo a las simulaciones efectuadas para la distribución exponencial con t.v.g., son los Casos 1 y 4 los que se presentan con mayor frecuencia. Es por ello que serán los únicos que se analizarán, ya que son suficientes para mostrar las dificultades analíticas que implica el utilizar el mismo procedimiento que se siguió para conocer la distribución del estadístico de la razón de verosimilitud, bajo la aproximación continua a la verosimilitud.

Con el fin de facilitar el análisis de la distribución del estadístico $-2 \log R_D(\theta_0; \mathbf{x})$, donde $\theta_0 = (\mu_0, \sigma_0)$, se considerará primeramente $\sigma = \sigma_0$ conocida y muestras de tamaño uno, para luego generalizar a muestras de tamaño n . Después se analizará la distribución de $-2 \log R(\theta_0; \mathbf{x})$ considerando ambos parámetros desconocidos y muestras de tamaño n .

3.5.2.1 Distribución de $-2 \log R_D(\theta_0; \mathbf{x})$ en muestras de tamaño uno y σ conocida

Para muestras de tamaño uno, se tiene que en la ecuación (3.5.3), $L_1(\theta; \mathbf{x})$ es creciente y $L_2(\theta; \mathbf{x})$ decreciente, luego el máximo se alcanza en el punto frontera del soporte de $L_1(\theta; \mathbf{x})$, y siempre se tiene que $\hat{\mu} = x_{(1)} - h$. Por tal razón, solamente se tendrían que analizar los Casos 1 y 2, de la Tabla 3.1, pero como de antemano se considerarían sólo los Casos 1 y 4, entonces se analizará solamente el Caso 1.

Caso 1. En este caso, tanto $\hat{\mu}$ como μ_0 se encuentran dentro del intervalo $(-\infty, x_{(1)} - h]$ y como

$$\begin{aligned} L_D(\mu, \sigma = \sigma_0; \mathbf{x}) &= L_1(\mu, \sigma = \sigma_0; \mathbf{x}) = P(x_{(1)} - h < X \leq x_{(1)} + h) \\ &= \exp\left(-\frac{x_{(1)}}{\sigma}\right) \exp\left(\frac{\mu}{\sigma}\right) \left[\exp\left(\frac{h}{\sigma}\right) - \exp\left(-\frac{h}{\sigma}\right) \right] = k_1 \exp\left(\frac{\mu}{\sigma}\right), \end{aligned}$$

donde $k_1 > 0$ y el emv se alcanza siempre en el extremo derecho del soporte de $L_1(\mu, \sigma; \mathbf{x})$, entonces se tiene que $\hat{\mu} = x_{(1)} - h$.

Como puede verse en el Apéndice B, Sección B.2.1,

$$-2 \log R_D(\mu_0; \mathbf{x}) = 2 \left(\frac{x_{(1)} - h - \mu_0}{\sigma_0} \right).$$

y

$$\begin{aligned} P(-2 \log R_D(\mu_0; \mathbf{x}) \leq z) &= P\left(x_{(1)} \leq \mu_0 + \frac{z\sigma_0}{2} + h\right) \\ &= 1 - \exp\left(-\frac{z}{2}\right) \cdot \exp\left(-\frac{h}{\sigma_0}\right) \end{aligned} \quad (3.5.6)$$

Usando nuevamente el argumento visto en la Sección 3.5.1, se concluye que $-2 \log R(\mu_0; \mathbf{x})$ sigue una distribución $\chi_{(2)}^2$, trasladada una distancia $2h/\sigma_0$, como se muestra en la Sección B.2.1 del Apéndice B.

3.5.2.2 Distribución de $-2 \log R_D(\mu_0; \mathbf{x})$ en muestras de tamaño n y σ conocida

Se considerarán ahora muestras de tamaño n , pero se mantiene el supuesto de que σ es conocida. De los casos presentados en la Tabla 3.1, nuevamente se estudian sólo los Casos 1 y 4.

Caso 1. En este caso se obtiene de nuevo, que $\hat{\mu} = x_{(1)} - h$ y si se considera $\sigma = \sigma_0$ conocida, la función de verosimilitud está dada por

$$\begin{aligned} L_D(\mu, \sigma = \sigma_0; \mathbf{x}) &= L_1(\mu, \sigma = \sigma_0; \mathbf{x}) = \prod_{i=1}^n P(x_{(i)} - h < X \leq x_{(i)} + h) \\ &= \exp\left(-\sum_{i=1}^n \frac{x_{(i)}}{\sigma_0}\right) \exp(n\mu/\sigma_0) \left[\exp(h/\sigma_0) - \exp(-h/\sigma_0) \right]^n, \end{aligned}$$

de donde se sigue que

$$\frac{L_1(\mu_0, \sigma = \sigma_0; \mathbf{x})}{L_1(\hat{\mu}, \sigma = \sigma_0; \mathbf{x})} = \frac{\exp(n\mu_0/\sigma_0)}{\exp[n(x_{(1)} - h)/\sigma_0]},$$

y repitiendo el procedimiento detallado en la Sección B.2.1, del Apéndice B, se llega a que la distribución de $-2 \log R(\mu_0; \mathbf{x})$ es la de una $\chi_{(2)}^2$, trasladada una distancia $2nh/\sigma_0$.

Caso 4. Este caso es una situación donde $\hat{\mu} \in (x_{(1)} - h, x_{(1)} + h]$ y $\mu_0 \in (x_{(1)} - h, x_{(1)} + h]$, por lo que $L_D(\mu, \sigma; \mathbf{x}) = L_2(\mu, \sigma; \mathbf{x})$, donde, suponiendo nuevamente σ conocida,

$$L_2(\mu, \sigma = \sigma_0; \mathbf{x}) = \left\{ 1 - \exp \left[-\frac{(x_{(1)} + h - \mu)}{\sigma_0} \right] \right\}^m \times \\ \exp \left(-\sum_{i=m+1}^n \frac{x_{(i)}}{\sigma_0} \right) \left[\exp \left(\frac{\mu + h}{\sigma_0} \right) - \exp \left(\frac{\mu - h}{\sigma_0} \right) \right]^{n-m} I_{(x_{(1)} - h, x_{(1)} + h]}(\mu),$$

a partir de la cual, como se muestra en la Sección B.2.2 del Apéndice B, puede obtenerse el estimador de máxima verosimilitud de μ , resultando en

$$\hat{\mu} = (x_{(1)} + h) + \sigma_0 \log \left(1 - \frac{m}{n} \right) \quad (3.5.7)$$

donde m es el número de veces que se repite la observación más pequeña.

Sin embargo, cuando se calcula

$$-2 \log \left[\frac{L_2(\mu_0, \sigma_0; \mathbf{x})}{L_2(\hat{\mu}, \sigma_0; \mathbf{x})} \right]$$

no es posible, usando este procedimiento, obtener una expresión sencilla para el estadístico de la razón de verosimilitud y conocer la distribución de éste. En el Apéndice B, Sección B.2.2, se muestra que considerando $p = P(\mu < X \leq x_{(1)} + h)$, puede llegarse a una versión empírica de la expresión (3.5.7).

3.5.2.3 Distribución de $-2 \log R_D(\mu_0; \mathbf{x})$ en muestras de tamaño n y σ desconocida

En las secciones anteriores se ha supuesto que se conoce el parámetro de escala σ de la distribución exponencial con tiempo de vida garantizado. Cuando no es así, es necesario estimarlo al igual que se estima el parámetro umbral. Sólo de esta manera se puede estudiar el comportamiento de $-2 \log R(\theta_0; \mathbf{x})$.

Caso 1. En este caso, se obtiene que el emv de μ es $\hat{\mu} = x_{(1)} - h$, y se llega a que el estimador de σ está dado por la expresión

$$\hat{\sigma} = 2h \cdot \left[\log \left(\frac{\bar{x} - \hat{\mu} + h}{\bar{x} - \hat{\mu} - h} \right) \right]^{-1}. \quad (3.5.8)$$

Caso 4. Para este caso, donde el emv de μ resulta

$$\hat{\mu} = (x_{(1)} + h) - \sigma \log \left(\frac{n}{n-m} \right),$$

también se obtiene que para el parámetro de escala σ , el estimador está dado por (3.5.8). Ambos desarrollos se muestran en la Sección B.2.3 del Apéndice B.

El estimador mostrado en (3.5.8), es muy similar al que resulta en el Ejemplo 1.2 y que Kalbfleisch (1985, p. 28) obtiene para el parámetro de escala en una distribución exponencial, cuando utiliza la verosimilitud exacta (1.3.12). La diferencia estriba en que (3.5.8) incluye al estimador $\hat{\mu}$. Es importante mencionar que también en los Casos 2 y 3, para los cuales no se incluyen los cálculos, se obtiene este mismo estimador de σ .

Cuando σ es desconocida, en ninguno de los casos expuestos en la Tabla 3.1 fue posible llegar a una expresión sencilla que permita explicar el comportamiento asintótico del estadístico $-2 \log R(\theta_0; \mathbf{x})$.

Por tanto, no es posible llegar a una expresión analítica simple que permita deducir la distribución del estadístico de la razón de verosimilitudes, cuando para esta densidad se utiliza la verosimilitud discretizada y se utiliza un procedimiento similar al aplicado con la aproximación continua. Mucho menos es posible lograrlo en un caso Weibull donde $\beta \neq 1$, ya que no existen estimadores explícitos para todos los parámetros del modelo.

3.5.3 Comparación entre las verosimilitudes restringida y discretizada

El modelo exponencial con t.v.g. es, como ya se dijo, un caso particular del modelo Weibull, cuando el parámetro de forma se fija en $\beta = 1$. Como pudo observarse en la sección anterior, se analizaron cuatro posibles escenarios para estudiar el comportamiento asintótico del estadístico de la razón de verosimilitudes y sólo bajo condiciones muy específicas pudo llegarse a una distribución conocida. Cuando este parámetro se deja variar libremente, ni la aproximación continua, ni la verosimilitud discretizada, permiten encontrar, analíticamente, todos los emv's de sus parámetros. Esto hace imposible deducir el comportamiento asintótico del estadístico de la razón de verosimilitud, siguiendo un análisis similar al efectuado con el uso de la aproximación continua.

Sin embargo, considerando que la verosimilitud discretizada, propuesta en esta tesis, es una especie de verosimilitud agrupada, similar a la que trabajan tanto Giesbrecht y Kempthorne (1976), como Hirose y Lai (1997), puede considerarse que los estimadores de los parámetros del modelo Weibull poseen las propiedades asintóticas que tienen los estimadores del caso multinomial. Este hallazgo nos permite afirmar que el comportamiento del estadístico de la razón de verosimilitudes para la verosi-

	n	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$FC(\sigma)$	# de I.V. a la izq. de σ	# de I.V. a la derecha de σ
$\mu = 8$	25	8695	0	1305	9469	407	124
	50	8778	0	1222	9493	343	164
$\sigma = 1$	100	8938	0	1062	9486	325	189
	500	9589	0	411	9511	264	225

Tabla 3.2: Datos simulados y analizados bajo un modelo exponencial con t.v.g, y utilizando la verosimilitud discretizada, con $2h = 0.01$.

	n	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$FC(\sigma)$	# de I.V. a la izq. de σ	# de I.V. a la derecha de σ
$\mu = 8$	25	8408	0	1592	9434	450	116
	50	8548	0	1452	9479	379	142
	100	8532	0	1468	9445	370	185
$\sigma = 1$	500	8479	0	1521	9498	275	227
	1000	8534	0	1466	9517	241	242

Tabla 3.3: Datos simulados y analizados bajo un modelo exponencial con t.v.g., y utilizando la aproximación continua a la verosimilitud

militud discretizada, es el mismo que en los casos regulares.

Para mostrar que computacionalmente se obtiene lo anteriormente dicho, se simularon diez mil muestras de una variable aleatoria discretizada X , con distribución exponencial con tiempo de vida garantizado, con tamaños 25, 50, 100 y 500, fijando los valores de los parámetros en $\mu = 8$ y $\sigma = 1$ y discretizando los datos con una precisión de $2h = 0.01$. La Tabla 3.2 muestra en las columnas $FC(\mu)$ y $FC(\sigma)$, las frecuencias de cobertura de los intervalos de verosimilitud de los parámetros μ y σ , calculados a un valor de $c = 0.1465$. Como puede verse en dicha tabla, el comportamiento del estadístico de la razón de verosimilitudes converge a una distribución $\chi^2_{(1)}$, tanto para el parámetro umbral como para el parámetro σ .

Por otra parte, se trabajaron también, diez mil muestras de tamaños 25, 50, 100, 500 y 1000, de una variable aleatoria Y con distribución exponencial con t.v.g. y con los mismos valores de parámetros, pero en su análisis se utilizó la aproximación continua. En la Tabla 3.3 puede observarse que la distribución asintótica de $-2 \log R(\mu_0; \mathbf{y})$, encontrada analíticamente, se cumple aún con muestras que no son necesariamente grandes, esto es, converge a una distribución $\chi^2_{(2)}$, resultado asintótico que demuestra Lawless (2003, p. 191). Es importante mencionar que los datos simulados para este análisis son los mismos que se utilizaron con el cálculo de la verosimilitud discretizada (aunque en ésta no se analizaron muestras de tamaño mil), pero sin efectuar en ellos el proceso de discretización, ya que al simular para $\theta_0 = (\mu_0, \sigma_0)$ y $2h$ fijo, y utilizar la aproximación continua a la verosimilitud, puede encontrarse un valor de n a partir del cual, el intervalo de verosimilitud calculado excluya el verdadero valor μ_0 del parámetro umbral.

En ambas tablas puede verse que los intervalos de verosimilitud perfil de μ que no cubren el verdadero valor del parámetro, lo sobreestiman. Sin embargo, en el caso del parámetro σ , la mayoría de los intervalos que no cubren el valor $\sigma = 1$, lo subestiman. Esta información da un indicio sobre la forma de la verosimilitud perfil del parámetro en estudio.

3.6 Simulaciones

En esta sección se comparan la verosimilitud discretizada y la verosimilitud restringida, con base en las inferencias que arrojan sobre los parámetros de un modelo no regular con singularidades, como el Weibull de tres parámetros. Resulta de particular interés estudiar el desempeño de dichas verosimilitudes en el caso de muestras pequeñas y de tamaño moderado. Esto se hará a través de las coberturas de los intervalos de verosimilitud perfil de los parámetros del modelo, bajo cada verosimilitud. Para ello, el uso de simulaciones es indispensable y crucial.

Se simularán diez mil muestras independientes de varios tamaños, con distribución Weibull de tres parámetros, bajo cuatro escenarios o conjuntos de parámetros distintos. Estos escenarios reflejan algunas de las situaciones prácticas, del mundo real, que se presentan como ejemplos en el Capítulo 4. Se considerarán también distintas precisiones del instrumento de medición, para simular las muestras discretizadas, esto con el fin de valorar el efecto que puedan tener éstas en las inferencias.

Uno de los resultados más importantes que se mostrarán en esta sección es que la verosimilitud discretizada para el modelo Weibull de tres parámetros, que es un modelo no regular, se comporta como si fuese la verosimilitud de un modelo regular, la de un modelo multinomial. Por tanto, la razón de verosimilitud discretizada tiene la misma distribución asintótica que la de casos regulares, si bien la velocidad de convergencia a esta distribución es más lenta para el caso de estimar un parámetro umbral. En aquellos casos que resultan problemáticos al usar la aproximación continua L_C , y que generalmente ocurren para datos provenientes de una Weibull con parámetro de forma $\beta < 1$, la verosimilitud discretizada requiere muestras que tengan al menos un tamaño $n \geq 500$ para que se cumplan los resultados asintóticos. Los otros parámetros del modelo no regular, se estiman razonablemente bien.

Por otra parte, la verosimilitud restringida también se comporta de manera similar y razonablemente bien en estos casos, como si se tratara de un modelo regular, aunque los intervalos de verosimilitud perfil son ligeramente más anchos que los que se obtienen con la verosimilitud discretizada. Sin embargo, la estimación del umbral no es tan precisa, puesto que el emv no se puede acercar tanto a la observación más pequeña, como puede hacerlo con la verosimilitud discretizada.

En resumen, se verá que las simulaciones muestran que tanto la verosimilitud discretizada propuesta como la verosimilitud restringida, siempre conducen a una

estimación eficiente y sin problemas numéricos, al estimar los parámetros en todos los escenarios que se consideraron.

En la siguiente subsección se describen los escenarios de simulación que se considerarán y las mediciones de interés que se presentarán para cada uno de ellos. Posteriormente se describen, con mayor detalle, los resultados de las simulaciones para cada uno de los escenarios. Finalmente se presenta un resumen de los resultados globales analizando todos los resultados obtenidos.

3.6.1 Selección de los escenarios de simulación

La selección de los escenarios de simulación pretende reflejar situaciones que pueden surgir en la realidad. En general, los valores de parámetros considerados en los escenarios fueron sugeridos por las estimaciones obtenidas para los datos de los ejemplos reales que se presentan en el Capítulo 4.

Tres de los escenarios de simulación ejemplifican las tres diferentes formas posibles que puede adoptar la verosimilitud perfil de un parámetro umbral cuando se usa la verosimilitud continua L_C y que se presentaron en la Sección 2.2. Se incluye también un cuarto escenario con el objetivo de comparar los resultados que Smith y Weissman (1985) obtuvieron para simulaciones con un modelo Weibull con $\beta = 1.25$. Este escenario puede ser problemático porque da lugar a una proporción alta de muestras con verosimilitudes con forma de J y sin máximo local, cuando se usa L_C . Los escenarios seleccionados se describen a continuación.

- **Escenario 1.** En este escenario se simulan muestras de una distribución Weibull con parámetros $\mu = 4$, $\sigma = 18$, $\beta = 0.7$, que son las estimaciones obtenidas en el Ejemplo 4.2.1 que dio lugar a una verosimilitud perfil del umbral en forma de J , como la mostrada en la Figura 2.1(b), al usar L_C . Por tanto, esto ocasionará que haya una proporción no despreciable de muestras simuladas bajo este escenario que den lugar a verosimilitudes con forma de J , cuando se utilice la aproximación continua.
- **Escenario 2.** En este escenario, los valores que se fijaron para los parámetros fueron $\mu = 8$, $\sigma = 6$ y $\beta = 2$. Estos reflejan, aproximadamente, los resultados que se obtienen en el Ejemplo 4.1.1, al modelar ciertos diámetros de troncos de árboles. En este ejemplo, la verosimilitud perfil del parámetro umbral tiene una forma como la presentada en la Figura 2.1(a), la cual presenta un máximo local no muy alejado de la singularidad.
- **Escenario 3.** Los valores de los parámetros que se utilizaron para simular las muestras correspondientes a este escenario fueron, $\mu = -3.5$, $\sigma = 5$, $\beta = 20$, que son aproximadamente los resultados obtenidos al ajustar una distribución Weibull de tres parámetros, a las mediciones de la fuerza de fibras de vidrio presentada por Smith y Naylor (1987a) y descrita en el Ejemplo 4.3.3. En

este caso la forma de la verosimilitud perfil del parámetro umbral es aplanada en una región del espacio parametral, como la mostrada en la Figura 2.1(c). Esta forma aplanada indica que un modelo Gumbel de dos parámetros también podría describir bien a los datos. Esta forma está en realidad asociada a un problema de selección de modelos.

- **Escenario 4.** En este último escenario los valores que se tomaron para los parámetros fueron $\mu = 0$, $\sigma = 1$, $\beta = 1.25$. Este es un escenario interesante pues por el valor asignado a β puede considerarse un caso intermedio entre los Escenarios 1 y 2. En este caso, las formas que toma la verosimilitud perfil del parámetro umbral son, por lo general, las presentadas en las Figuras 2.1(a) y 2.1(b).

3.6.2 Proceso de discretización de datos simulados

Para cada uno de los cuatro escenarios anteriores se simularon diez mil muestras Weibull de tamaños $n = 25$, 50 y 100. Se discretizó cada una de las observaciones de estas muestras, con el fin de simular el uso de un instrumento de medición con una precisión de $2h$. Las muestras que se obtuvieron, fueron resultado de aplicar el proceso de discretización descrito en la Sección 3.2.

En todos los escenarios se utilizaron tres precisiones diferentes para discretizar las observaciones simuladas. Gracias a ello fue posible estudiar el efecto que la precisión de un instrumento puede tener en las coberturas de los intervalos de verosimilitud que se describirán en la Sección 3.6.3. Una de las precisiones utilizadas corresponde a la precisión del instrumento de medición con el que se midieron las observaciones en el ejemplo que se tomó de base para fijar los valores de los parámetros del escenario; ese valor de $2h$ aparece con negritas en la Tabla 3.4. Para elegir las otras dos precisiones se tuvo cuidado de que la partición generada por la precisión $2h$, no diera lugar a muchas observaciones repetidas, pues un instrumento de medición (para variables aleatorias continuas) se considera bueno en la medida que no haya mediciones repetidas.

La quinta columna de la Tabla 3.4, etiquetada como $\max P_i$, muestra la probabilidad máxima que alcanzan los intervalos de esta partición, bajo el modelo teórico considerado. Así, para controlar el número de observaciones repetidas se verificó que los intervalos generados por la partición, que induce el instrumento de medición simulado, tuvieran una probabilidad máxima asociada menor de 0.02, dado los tamaños de muestra que se utilizarían. De esta manera, en muestras de tamaño 100 se espera no tener más de dos observaciones repetidas, como se muestra en la sexta columna, etiquetada como $100 \max P_i$, de esta misma tabla. La Tabla 3.4 también incluye el valor esperado y desviación estándar de la variable aleatoria Weibull considerada. Estas medidas sirvieron para escoger los límites de la partición que se usaría para llevar a cabo el proceso de discretización. Esto es importante para lograr cálculos computacionales más eficientes.

Valores de los parámetros	$E(X)$	$\sqrt{V(X)}$	$2h$	$\max P_i$	$100 \times \max P_i$
Escenario 1 Weibull: $[\mu = 4, \sigma = 18, \beta = 0.7]$ DGVE: $[a = -22, b = 25.71, c = -1.43]$	26.7848	33.3211	0.01 0.005 0.001	0.0053 0.0032 0.0010	0.5250 0.3235 0.1050
Escenario 2 Weibull: $[\mu = 8, \sigma = 6, \beta = 2]$ DGVE: $[a = -14, b = 3, c = -0.5]$	13.3174	2.7795	0.1 0.05 0.01	0.0143 0.0071 0.0014	1.4295 0.7148 0.1430
Escenario 3 Weibull: $[\mu = -3.5, \sigma = 5, \beta = 20]$ DGVE: $[a = -1.5, b = 0.25, c = -0.05]$	1.3675	0.3017	0.01 0.005 0.001	0.0147 0.0074 0.0015	1.4732 0.7367 0.1473
Escenario 4 Weibull: $[\mu = 0, \sigma = 1, \beta = 1.25]$ DGVE: $[a = -1, b = 0.8, c = -0.8]$	0.9314	0.7498	0.02 0.01 0.005	0.0148 0.0074 0.0037	1.4833 0.7417 0.3709

Tabla 3.4: Escenarios considerados en las simulaciones

3.6.3 Cantidades de interés en el estudio de simulación

En esta sección se describen las cantidades de interés que son medidas en el estudio de simulación y cuyos resultados se presentan en la Sección 3.6.4. Estas cantidades se calcularon tanto para la verosimilitud discretizada, como para la verosimilitud restringida. La observación más pequeña de la muestra discretizada se denota con $x_{(1)}$, la precisión del instrumento de medición se representa con $2h$ y $FC(\cdot)$ denota la frecuencia de cobertura del intervalo de verosimilitud perfil del parámetro indicado en el paréntesis.

Bajo ambas verosimilitudes se calcularon las frecuencias de cobertura de los intervalos de verosimilitud perfil del parámetro de interés; esto es, el número de intervalos que capturan el verdadero valor del parámetro. La proporción de estos intervalos aproxima la probabilidad teórica de cobertura de los mismos y permite valorar a partir de qué tamaño de muestra se cumplen los resultados asintóticos previstos para casos regulares. Algo importante de señalar es que en la literatura estadística que se revisó, no se han descrito las frecuencias de cobertura de intervalos de verosimilitud perfil restringida para modelos no regulares, sobre todo en el caso de muestras moderadas y pequeñas.

Por otra parte, se cuantificaron además, los intervalos que subestiman o sobrestiman el parámetro de interés. La proporción de intervalos de verosimilitud perfil que subestima o bien que sobrestima al verdadero valor del parámetro, depende de si la asimetría de las funciones de verosimilitud correspondientes es muy marcada. En casos regulares con tamaños de muestra adecuados, las proporciones de intervalos que no cubren el verdadero valor del parámetro suelen, por lo general, ser similares.

Todas las frecuencias de cobertura se calcularon utilizando intervalos de verosimilitud perfil de un nivel de verosimilitud de $c = 0.1465$. Se seleccionó este valor de c , pues en casos regulares está asociado a intervalos de confianza del 95%, en los

que el estadístico de la razón de verosimilitudes se distribuye como una $\chi^2_{(1)}$. En todos los escenarios y para los diferentes tamaños de muestras, se analiza el efecto que pudiera tener la precisión del instrumento de medición en las coberturas de los intervalos de verosimilitud.

Así pues, las cantidades de interés a que se presentan en las distintas tablas, como resultado de estas simulaciones son las siguientes:

- **Frecuencias de cobertura de los intervalos de verosimilitud perfil de un parámetro de interés.**

Se calcularon las frecuencias de cobertura de los intervalos de verosimilitud perfil de los parámetros μ , σ y β del modelo Weibull. En las tablas que se presentan, dichas coberturas aparecen en las columnas cuyo encabezado es $FC(\mu)$, $FC(\sigma)$ y $FC(\beta)$, respectivamente.

Se calculó también la frecuencia de cobertura de los intervalos de verosimilitud perfil del cuantil teórico $Q_{0.05}$, de una distribución Weibull con parámetros μ , σ y β . La finalidad de ello es, como se ha descrito a través de esta tesis, la ventaja que puede tener el estimar un cuantil pequeño cuando se dificulta la estimación de un parámetro umbral. Ahora, sólo para el Escenario 3, se calculó también la frecuencia de cobertura de los intervalos de verosimilitud perfil del cuantil teórico $Q_{0.05}$ de un modelo Gumbel, ya que en este escenario los valores de los parámetros que se utilizaron para simular las muestras Weibull, pueden originar muestras donde un modelo Gumbel sea adecuado a los datos.

Es importante mencionar que Bolívar (2007) calcula la frecuencia de cobertura de los intervalos de verosimilitud perfil de cuantiles grandes, para distribuciones de máximos y afirma que proporcionan inferencias más razonables y mejores coberturas, en muestras de tamaño moderado, que las obtenidas con los intervalos de confianza asintóticos de máxima verosimilitud. La razón es que dichos intervalos asintóticos pueden subestimar o sobrestimar el cuantil de interés si las muestras no son suficientemente grandes.

Se presentan también algunas tablas donde se analiza la información que proporciona el parámetro de forma c de la DGVE, tanto con la verosimilitud discretizada como con la restringida. Se contabilizó el número de intervalos de verosimilitud que cubren el valor $c = 0$. Estas frecuencias aparecen en la columna $R_p(c = 0) > 0.1465$. La finalidad de ello es ratificar si las muestras indican de manera clara que fueron simuladas bajo un modelo Weibull o no. Por otra parte, también permiten ver aquellos casos donde un modelo Gumbel, asociado a $c = 0$, pudiera explicar bien los datos; estos casos son aquellos donde el intervalo de c incluye al cero. Como las muestras se simularon bajo un modelo Weibull, se esperarí que la mayoría de estos intervalos tengan ambos extremos negativos. Sin embargo en el Escenario 3, el modelo Weibull

correspondiente está muy cercano a su modelo límite Gumbel, por lo que se esperaba que ambos modelos pudieran describir bien a estas muestras.

- **Número de intervalos que sobreestiman o subestiman cierto parámetro**

En el caso del parámetro umbral, aparte de calcular la frecuencia de cobertura de sus intervalos de verosimilitud, se caracterizaron los intervalos que no cubrieron el valor de μ con el cual se simularon las muestras. Así las columnas *# de I.V. a la derecha de μ* y *# de I.V. a la izq. de μ* contabilizan los intervalos que sobreestiman o subestiman el verdadero valor de μ . Es importante contabilizar estos intervalos pues, dependiendo del contexto del problema, una subestimación o bien sobrestimación, puede llevar a conclusiones equivocadas y ser sumamente peligrosa en casos de prevención de catástrofes.

Para el parámetro de forma c de la DGVE, aquellos intervalos donde $R_p(c = 0) < 0.1465$ son los que no cubren el valor $c = 0$. Estos intervalos se dividieron en los que se ubican a la izquierda de $c = 0$ ó bien a la derecha de este valor. El parámetro c de la DGVE es muy informativo en cuanto a la selección de un modelo de valor extremo. Recuérdese que si $c < 0$ se trata de un modelo Weibull, $c = 0$ es Gumbel y si $c > 0$ se trata de un modelo Fréchet. Cuando el intervalo de verosimilitud perfil de c incluye al cero, la práctica usual en Teoría de Extremos es usar la DGVE para calcular la verosimilitud perfil de los cuantiles de interés. La idea es no restringir el modelo a una sola familia de las tres posibles de valores extremos, aceptando la incertidumbre que hay en los datos sobre el verdadero modelo de atracción para los extremos. Esto se considera una medida conservadora.

Por otra parte, las muestras cuyos intervalos excluyeron al cuantil verdadero donde $R_p(Q_{0.05}) < 0.1465$, se contabilizan en dos columnas que dividen estas muestras en las que $\hat{Q}_{0.05} < Q_{0.05}$ (*# de I.V. a la izq. de $Q_{0.05}$*) y aquellas donde $\hat{Q}_{0.05} > Q_{0.05}$ (*# de I.V. a la derecha de $Q_{0.05}$*). Para cada uno de los escenarios analizados, se comparan las coberturas obtenidas con los intervalos de verosimilitud perfil del cuantil $Q_{0.05}$ y del umbral μ .

- **Características de los emv de β y de μ .**

En todos los escenarios también se contabilizó si $\hat{\beta} \geq 1$ o bien, $\hat{\beta} < 1$, ya que esta información ayuda a caracterizar la forma de la densidad Weibull que pueden seguir los datos, así como la forma de la verosimilitud perfil de μ . Cuando $\hat{\beta} < 1$, se trata de una situación donde estimar el parámetro umbral es mucho más difícil y se van a necesitar muestras de mayor tamaño. Para todas las muestras donde $\hat{\beta} < 1$, la verosimilitud perfil restringida del umbral y la obtenida de la aproximación continua tendrán forma de J . Cuando se usa L_C , estos casos se identifican como problemáticos y se dice que no existe el emv; sin embargo cuando en estas muestras se usa la verosimilitud discretizada, siempre se encuentra el emv y un intervalo de estimación, sin problema alguno. Hay que notar que también puede darse una forma de J en algunos casos donde el

emv de β es cercano a uno e inclusive $\hat{\beta} \geq 1$.

Por otra parte, el parámetro umbral se caracterizó de acuerdo a la ubicación que tiene con respecto al intervalo que rodea a la observación discretizada más pequeña, $x_{(1)}$. Cuando se utiliza la verosimilitud discretizada se consideró la posibilidad de que el emv $\hat{\mu}$ estuviera dentro del intervalo que rodea a $x_{(1)}$. Así pues, para la verosimilitud discretizada, se contabilizó el número de veces que $\hat{\mu}$ cae en alguno de los siguientes tres intervalos posibles: $(-\infty, x_1 - h)$, $[x_{(1)} - h, x_{(1)})$ y $[x_{(1)}, x_{(1)} + h]$. Con respecto a la verosimilitud restringida, ésta solamente se evalúa hasta $x_1 - h$, por lo que la ubicación de $\hat{\mu}$ se caracterizó como $\hat{\mu} < x_{(1)} - h$ o bien $\hat{\mu} = x_{(1)} - h$. Para contabilizar las frecuencias anteriores, computacionalmente se consideró que dos números son iguales si difieren en menos de una milésima.

3.6.4 Resultados de las simulaciones

En esta sección se presentan varias tablas que resumen los resultados que se derivan del análisis de las muestras simuladas para cada uno de los escenarios antes descritos. Para todos se presenta una tabla con las cantidades de interés que se han descrito, usando tanto la verosimilitud discretizada como la restringida.

3.6.4.1 Escenario 1

[Weibull: $\mu = 4$ y 4.003 , $\beta = 0.7$, $\sigma = 18$; DGVE: $a = -22$, $b = 25.71$, $c = -0.5$]

Este es el escenario de mayor dificultad para estimar el parámetro umbral de la distribución Weibull. Como $\beta < 1$, muchos emv $\hat{\beta}$ también serán menores a uno y esto ocasiona que la aproximación continua a la verosimilitud perfil del umbral tenga forma de J y sea no acotada. En contraste, la verosimilitud restringida es acotada, si bien también tiene forma de J. La verosimilitud discretizada también es acotada y generalmente alcanza el máximo en el intervalo que rodea a la observación más pequeña. Los intervalos de verosimilitud perfil del umbral son similares para estas dos últimas verosimilitudes. Sin embargo, se necesita que la muestra sea grande, $n = 500$, para tener coberturas razonables. En contraste para los parámetros restantes, los intervalos de verosimilitud perfil bajo ambas verosimilitudes tienen buenas coberturas incluso para muestras chicas con $n = 25$.

En este escenario se considera la posibilidad de que el parámetro umbral tome el valor del límite inferior de uno de los intervalos de la partición que induce el instrumento de medición (cuando $\mu = 4$), llamado Escenario 1a, y también se considera la posibilidad de que este parámetro se encuentre dentro de uno de estos intervalos (cuando $\mu = 4.0073$), y es el llamado Escenario 1b. En el primer caso se utilizaron los valores 0.01, 0.005 y 0.001 como las precisiones simuladas para el instrumento de medición y en el segundo se consideró tan sólo una precisión de 0.01, que es la

observada en el ejemplo que se tomó de base para fijar los valores de los parámetros.

Los resultados que se obtienen al usar las verosimilitudes discretizada y restringida, en este escenario, se presentan en las Tablas 3.5, 3.6, 3.7, 3.8 y 3.9.

- **Frecuencia de cobertura de los intervalos de verosimilitud para los parámetros de un modelo Weibull.**

En la Tabla 3.5 puede observarse que, cuando se utiliza la verosimilitud discretizada, la cobertura empírica de los intervalos de verosimilitud perfil del parámetro umbral $\mu = 4$ se hace más pequeña conforme $2h$ disminuye. Sin embargo, la cobertura aumenta conforme crece el tamaño de muestra. Con el fin de estudiar el comportamiento de estas coberturas en una muestra más grande, se simularon diez mil muestras de tamaño 500 para las cuales se logra una cobertura de aproximadamente 0.93 cuando se fija la precisión en $2h = 0.01$.

Los resultados que arroja la verosimilitud restringida para el Escenario 1, son muy similares y se muestran en la Tabla 3.6. En este caso, los intervalos de verosimilitud perfil del umbral llegan a alcanzar una cobertura aproximada de 0.83 cuando se tienen muestras de tamaño cien y una precisión de $2h = 0.01$ y al simular muestras de tamaño 500 se alcanza una cobertura de casi 0.95. La coberturas del umbral, obtenidas con la verosimilitud restringida, parecen un poco mejores que las obtenidas con la verosimilitud discretizada, pero esto se debe a que los intervalos que resultan con la verosimilitud restringida son más anchos que los que arroja la verosimilitud discretizada, cuando el parámetro de forma $\beta < 1$. La razón de ello es que la verosimilitud relativa restringida para μ se estandariza con un valor menor que el utilizado en la verosimilitud relativa discretizada de este parámetro umbral. Esto se puede apreciar en la verosimilitud perfil del umbral del Ejemplo 4.2.1.

Los resultados que se obtienen, tanto con la verosimilitud discretizada como con la restringida, cuando se considera un umbral $\mu = 4.0073$ son muy similares a los descritos anteriormente, como puede constatarse en las Tablas 3.7 y 3.8.

Aún cuando el Escenario 1 es una situación generalmente problemática, las coberturas para el parámetro de escala σ y el de forma β , son bastante buenas y de aproximadamente 0.95 cuando se usa la verosimilitud discretizada y un poco mejores con la verosimilitud restringida.

- **Frecuencia de cobertura de los intervalos de verosimilitud perfil de un cuantil de interés.**

Las coberturas empíricas de los intervalos de verosimilitud perfil de un cuantil como $Q_{0.05}$, mostradas en la columna $FC(Q_{0.05})$, son mucho mejores que las

de un parámetro umbral, pero aún así se requieren muestras de 50 o mayores, para tener coberturas más o menos razonables. Esto ocurre tanto en el Escenario 1a, como el 1b. En la Tabla 3.5, puede observarse que estas coberturas empíricas disminuyen conforme disminuye el valor $2h$ asociado a la precisión considerada para el instrumento de medición. Sin embargo, se mantienen estables para muestras de tamaño 50 y 100, a pesar de variar la precisión del instrumento.

Estas frecuencias de cobertura mejoran cuando el cuantil no es tan pequeño. Por ejemplo, para el caso de la verosimilitud discretizada se calcularon las coberturas empíricas de los intervalos de verosimilitud perfil de los cuantiles $Q_{0.10}$ y de la mediana $Q_{0.50}$. Para $Q_{0.10}$ las coberturas obtenidas son razonables a partir de muestras de tamaño 25, pero, para $Q_{0.50}$ se cumplen los resultados asintóticos aún con muestras de tamaño 25. Estos resultados se muestran también en la Tabla 3.5.

- **Número de intervalos que sobreestiman o subestiman cierto parámetro.**

Todos los intervalos que no incluyen al verdadero parámetro umbral μ caen a la derecha de éste, tanto con la verosimilitud discretizada, como con la restringida. El número de estos intervalos aumenta cuando $2h$ disminuye. En estos casos lo que sucede es que la observación $x_{(1)}$ no se encuentra cercana a μ y por ello el intervalo y el emv $\hat{\mu}$ sobreestiman a este parámetro. La forma de la verosimilitud perfil del umbral, en este escenario, es por lo general, muy asimétrica y sesgada a la izquierda. De manera contraria, los intervalos de verosimilitud perfil que no cubren al parámetro σ o bien, al parámetro β , por lo general lo subestiman. La forma de estas perfiles es también asimétrica y ligeramente sesgada a la derecha. Su asimetría es menos marcada que para la perfil del umbral.

Con respecto a la estimación de ciertos cuantiles, la mayoría de los intervalos que no cubren el cuantil $Q_{0.05}$, generalmente lo sobreestiman. Este comportamiento es más marcado cuando se utiliza la verosimilitud discretizada, como puede verse en las Tablas 3.5 y 3.6 para el Escenario 1a, donde se simuló con $\mu = 4$ y en las Tablas 3.7 y 3.8, para el Escenario 1b, en donde se consideró $\mu = 4.0073$. La verosimilitud perfil de $Q_{0.10}$ no es tan asimétrica como la de $Q_{0.05}$ y los intervalos que no cubren el valor teórico de $Q_{0.10}$ se encuentran tanto a la derecha como a la izquierda de éste. Lo mismo puede observarse para la mediana, $Q_{0.50}$, donde el número de intervalos que no cubren al verdadero valor se encuentran tanto a su derecha como a su izquierda, con frecuencias muy similares. Esto puede verse en la Tabla 3.5.

- **Características de los emv de μ y de β .**

Algo importante de notar en la Tabla 3.5, que resume los resultados de la ve-

rosimilitud discretizada, es que se tienen casos donde la estimación del umbral es mayor o igual que la observación más pequeña, lo cual se contabiliza en la columna $\hat{\mu} \in [x_{(1)}, x_{(1)} + h]$. Esta es una posibilidad que en la literatura no se ha considerado y que nunca ocurre cuando se utiliza la aproximación continua a la verosimilitud o bien la verosimilitud restringida. Estos casos ocurren con mayor frecuencia cuando se considera que el verdadero valor del umbral se encuentra dentro de uno de los intervalos resultantes de la partición, como puede constatararse en la Tabla 3.7, donde se simularon muestras con $\mu = 4.0073$. Los casos donde la estimación del umbral es igual a $x_{(1)}$ o mayor, son casos problemáticos cuando se utiliza la aproximación continua a la verosimilitud. Cuando se utiliza la verosimilitud restringida, la estimación numérica del parámetro umbral siempre es menor o igual que $x_{(1)} - h$, como puede verse en la Tabla 3.6, donde se describe la ubicación de $\hat{\mu}$.

Con respecto al parámetro de forma β , tanto con la verosimilitud restringida como con la discretizada, se tiene que más del 99% de las estimaciones de este parámetro son menores que uno, lo cual concuerda con el valor de β con el cual se simularon los datos.

La proporción de muestras donde la verosimilitud restringida arroja la estimación $\hat{\mu} = x_{(1)} - h$, es una cota superior de la proporción de muestras que resultarían problemáticas al utilizar L_C , ya que puede haber muestras donde la aproximación continua encuentre un máximo local en el intervalo $(x_{(1)} - h, x_{(1)})$, que la restringida no podría hallar por ser evaluada hasta $x_{(1)} - h$.

- **Frecuencia de cobertura de los intervalos de verosimilitud para el parámetro de forma c de la DGVE.**

Las muestras simuladas bajo el modelo Weibull seleccionado, también se analizaron con la DGVE y sólo en muestras de tamaño 25 se observó que a lo más tres de estos intervalos cubrieron el valor $c = 0$, cuando se utilizó la verosimilitud discretizada. Esto no ocurrió en ninguna de las muestras analizadas con la verosimilitud restringida. Es decir los intervalos de verosimilitud perfil identificaron correctamente, en más del 99% de las muestras, que el modelo que las generó era un Weibull (con parámetro de forma de la DGVE $c = -1.42$) y no uno Gumbel (con $c = 0$). Esto puede observarse en la Tabla 3.9.

3.6.4.2 Escenario 2

[Weibull: $\mu = 8, \sigma = 6, \beta = 2$, DGVE: $a = -14, b = 3, c = -0.5$]

Para este escenario el parámetro de forma Weibull es mayor a uno, pues se simularon muestras con $\beta = 2$. Con ello casi todos los env $\hat{\beta}$ serán mayores que uno, por lo que la verosimilitud continua L_C casi siempre tiene un máximo local,

aparte de la singularidad. Este máximo local coincide con los máximos globales de las verosimilitudes restringida y discretizada. Por ello, este escenario denota un caso prácticamente regular, como se observa en las coberturas de los intervalos de verosimilitud perfil del parámetro umbral y de los parámetros restantes. Los resultados obtenidos al analizar estas muestras con la verosimilitud discretizada y la restringida se resumen en las Tablas 3.10, 3.11 y 3.12.

- **Frecuencia de cobertura de los intervalos de verosimilitud para los parámetros de un modelo Weibull.**

En este escenario, los intervalos de verosimilitud perfil del parámetro umbral μ arrojan una cobertura empírica muy buena, con valores entre 0.91 y 0.94, incluso para muestras pequeñas, tanto cuando se utiliza la verosimilitud restringida como la discretizada. Estas coberturas aumentan conforme crece el tamaño de muestra, sin que haya diferencias notables con respecto a cambios en la precisión simulada para el instrumento de medición.

Con respecto a los parámetros σ y β , las coberturas de sus intervalos de verosimilitud perfil son bastante buenas, pues toman valores entre 0.91 y aproximadamente 0.95, tanto con la verosimilitud restringida como con la discretizada. Estas coberturas también aumentan conforme crece el tamaño de la muestra. En el caso de la verosimilitud restringida, para un tamaño de muestra fijo, la cobertura de los intervalos de verosimilitud para los parámetros del modelo Weibull aumentan conforme el instrumento de medición simulado es más preciso.

- **Frecuencia de cobertura de los intervalos de verosimilitud para un cuantil de interés.**

Las coberturas empíricas que se obtuvieron con los intervalos de verosimilitud perfil del cuantil $Q_{0.05}$ fueron bastante buenas ya que oscilan entre 0.91 y 0.94, tanto con la verosimilitud discretizada como con la restringida. Estas coberturas son muy similares a las obtenidas para el parámetro umbral, ya que este escenario no resulta problemático en cuanto a la estimación de los parámetros del modelo Weibull.

- **Número de intervalos que sobreestiman o subestiman cierto parámetro.**

Con ambas verosimilitudes se observa que la mayoría de los intervalos de verosimilitud que no cubren el valor teórico del umbral, se ubican a la derecha de éste. Esto también ocurre cuando se analizan los intervalos de verosimilitud perfil para el cuantil $Q_{0.05}$. Por el contrario, en la Tabla 3.10 puede verse que los intervalos de verosimilitud que no cubren los valores teóricos de σ y β , se ubican a la izquierda de éstos. Este comportamiento es similar al observado en el Escenario 1, sólo que en ese escenario la proporción de intervalos que no cubrían al verdadero valor del parámetro era ligeramente menor.

- **Características de los emv de μ y de β .**

La ubicación del emv del umbral con respecto a la observación más pequeña,

cuando se usa la verosimilitud discretizada, arroja que más del 95% de estas estimaciones fueron tales que $\hat{\mu} < x_{(1)} - h$. No se presenta ningún caso donde $\hat{\mu}$ tome valores en el intervalo $[x_{(1)} - h, x_{(1)} + h]$, ya que esta situación parece darse cuando los datos simulados provienen de una densidad Weibull convexa donde el parámetro de forma es $\beta \leq 1$, o bien mayor pero cercano a uno. Cuando se usa la verosimilitud restringida, el porcentaje de estimaciones del umbral que satisfacen $\hat{\mu} < x_{(1)} - h$ inicia en un 91% para muestras de tamaño 25, llegando al 100% para muestras de tamaño 100 analizadas con una precisión de $2h = 0.01$.

Por otra parte, con ambas verosimilitudes, más del 96% de las estimaciones del parámetro de forma son tales que $\hat{\beta} \geq 1$. Esto concuerda con el parámetro de forma del modelo con el cual se simularon las muestras. En este escenario, solamente alrededor del 3% de las muestras hubieran tenido problemas de estimación al usar la verosimilitud continua, con la que hubiesen tenido forma de J , sin máximo local.

- **Frecuencias del cobertura de los intervalos de verosimilitud para el parámetro de forma c de la DGVE.**

Tanto con la verosimilitud discretizada como con la restringida, en simulaciones efectuadas con muestras con $n = 25$, se observa que el 87% de los intervalos de verosimilitud perfil del parámetro de forma c caen por debajo de $c = 0$, corroborando que un modelo Weibull es adecuado. Cerca del 13% de estos intervalos cubren el valor $c = 0$, indicando que tanto un Weibull como un modelo Gumbel podrían ser adecuados. Sin embargo, cuando el tamaño de muestra se aumenta a 50 este porcentaje se reduce a menos del 1%, y es cero cuando las muestras simuladas son de tamaño 100. Estos resultados pueden observarse en la Tabla 3.12.

3.6.4.3 Escenario 3

Weibull: $\mu = -3.5, \sigma = 5, \beta = 20$, **DGVE:** $a = -1.5, b = .25, c = -0.05]$

Este escenario muestra situaciones donde hay dos modelos razonables a los datos, Weibull y Gumbel. Esto se refleja en la forma de las perfiles de los parámetros del modelo Weibull, pues son muy asimétricas y aplanadas. Por este motivo, varios autores han criticado a la verosimilitud perfil del umbral, calificándola de problemática y no informativa. Sin embargo, en esta tesis se plantea que precisamente la forma aplanada de la verosimilitud perfil indica que hay un modelo adicional más sencillo que puede describir bien a los datos, y es por tanto, muy informativa. En este escenario los valores de los parámetros se basan en las estimaciones de los datos de fibra de vidrio que analizaron Smith y Naylor (1987b) y que se presentan en la Sección 4.3.3 del Capítulo 4.

Bajo este escenario, se observa que son necesarias muestras mucho muy grandes para poder distinguir al modelo Weibull de su modelo límite Gumbel, del cual está

cercano. El modelo límite Gumbel tiene como parámetros, en la DGVE, los valores $[a = -1.5, b = .25, c = 0]$. Para fines prácticos, este modelo y el modelo bajo el cual se simularon los datos, son casi indistinguibles en el rango de valores que toman las muestras. Esto puede observarse cuando se construye una curva ROC para estas distribuciones, en la que se grafican los puntos cuyas coordenadas son estas dos distribuciones asociadas, $[F(z), G(z)]$, para una rejilla de valores z asociados a probabilidades no despreciables en estos modelos. En ella lo que se obtiene es una curva prácticamente indistinguible de la recta a 45 grados, indicando la similitud entre las dos distribuciones.

En este tipo de modelos puede interesar estimar una cota inferior para los datos, por lo que se decidió estimar un cuantil pequeño como el $Q_{0.05}$ bajo ambos modelos Weibull y Gumbel, y comparar estas estimaciones. Para ello que se contabilizan las coberturas que alcanzan los intervalos de verosimilitud perfil del cuantil $Q_{0.05}$, contruidos tanto bajo un modelo Gumbel como un Weibull. Todo el análisis que se describe para este escenario, se deduce de los resultados presentados en las Tablas 3.13, 3.14 y 3.15.

- **Frecuencia de cobertura de los intervalos de verosimilitud para los parámetros de un modelo Weibull.**

Cuando se utiliza la verosimilitud discretizada, como puede verse en la Tabla 3.13, las coberturas empíricas que alcanzan los intervalos de verosimilitud perfil del parámetro umbral fluctúan entre 0.96 y aproximadamente 0.97, sin que parezca afectar demasiado la precisión del instrumento de medición. Estas coberturas aumentan ligeramente conforme aumenta el tamaño de muestra. El comportamiento es muy similar cuando se usa la verosimilitud restringida, como se muestra en la Tabla 3.14.

Coberturas similares a las anteriormente mencionadas, se alcanzan con los intervalos de verosimilitud perfil para los parámetros σ y β . Esto ocurre con ambas verosimilitudes, sin que parezca influir, nuevamente, la precisión del instrumento de medición.

- **Frecuencia de cobertura de los intervalos de verosimilitud para el cuantil $Q_{0.05}$.**

Con la verosimilitud discretizada, las coberturas empíricas que alcanzan los intervalos de verosimilitud perfil del cuantil $Q_{0.05}$, contruidos con un modelo Weibull, oscilan entre 0.89 y 0.94 y son bastante estables dentro un mismo tamaño de muestra. Analizando las mismas muestras generadas con este modelo Weibull, pero estimando ahora un modelo Gumbel, las coberturas de los intervalos de verosimilitud perfil son mucho mejores desde muestras pequeñas y moderadas (entre 0.94 y 0.95).

Cuando se utiliza la verosimilitud restringida, las coberturas empíricas que alcanzan los intervalos de verosimilitud perfil del cuantil $Q_{0.05}$ bajo un modelo

Gumbel, son mejores que la obtenidas con un modelo Weibull, en las muestras de tamaños 25 y 50. Sin embargo, esto no se cumple en las muestras de tamaño 100, donde las coberturas obtenidas bajo el modelo Weibull son mejores que las logradas bajo el modelo Gumbel. La explicación es que para muestras más grandes, se puede discriminar mejor y notar que el modelo verdadero es un Weibull, cercano a su límite Gumbel, pero sin alcanzarlo.

- **Número de intervalos que sobreestiman o subestiman cierto parámetro.**

Con ambas verosimilitudes, la mayoría de los intervalos que no cubren el parámetro umbral, se encuentran a la derecha de éste, como ocurre también en los Escenarios 1 y 2. Con respecto a los intervalos de verosimilitud que no cubren el cuantil $Q_{0.05}$, cuando se utiliza un modelo Weibull, la mayoría de éstos se encuentran a la derecha de este cuantil. Existen algunos que subestiman este parámetro, cuando se utiliza la verosimilitud discretizada, pero con la verosimilitud restringida, casi la totalidad cae a la derecha de este cuantil. Bajo un modelo Gumbel, todos los intervalos que no cubren este parámetro, se encuentran a la derecha del mismo, tanto con la verosimilitud restringida como con la discretizada. Por último, los intervalos de verosimilitud perfil de σ y de β que excluyen al verdadero valor del parámetro, siempre lo subestiman y caen por abajo de él.

- **Características de los emv de μ y de β .**

Con respecto al parámetro de forma β , el 99.9% de las muestras identificaron que $\hat{\beta} > 1$. Por tanto estas muestras tienen, con la aproximación continua L_C , un máximo local que coincide con los máximos globales de la verosimilitud discretizada y restringida. Con estas dos verosimilitudes, la totalidad de las estimaciones del umbral se ubicaron a la izquierda de $x_{(1)} - h$, y favorecieron valores muy grandes negativos del umbral, al ser la perfil de forma muy aplanada en gran parte del espacio parametral.

- **Frecuencias del cobertura de los intervalos de verosimilitud para el parámetro de forma c de la DGVE.**

Como puede observarse en la Tabla 3.15, cuando se modela con la DGVE y se utiliza la verosimilitud discretizada, entre el 89% y el 93% de los intervalos de verosimilitud perfil de c cubren el valor de $c = 0$, lo cual es indicativo que un modelo Gumbel también es adecuado para modelar los datos, sobre todo en el caso de muestras pequeñas y moderadas. Los porcentajes mayores ocurren cuando el tamaño de muestra es más pequeño. Los intervalos que no incluyen al cero siempre están por abajo del mismo, sustentando un modelo Weibull.

Con la verosimilitud restringida, los intervalos de verosimilitud perfil de c siempre cubren el valor $c = 0$, para cualquier tamaño de muestra y precisión utilizada. Esto ocurre porque son ligeramente más anchos que los intervalos que se obtienen con la verosimilitud discretizada. Con la verosimilitud restringida

se identifica siempre que un modelo Gumbel es razonable para modelar estos datos y nunca se elige como opción única al modelo Weibull.

3.6.4.4 Escenario 4

Weibull [$\mu = 0, \sigma = 1, \beta = 1.25$], **DGVE**: $a = -1, b = 0.8, c = -0.8$

La razón de considerar este escenario es para poder comparar los resultados de las verosimilitudes discretizada y restringida, con los obtenidos por Smith y Weissman (1985) quienes consideran los mismos valores de parámetros en sus simulaciones, pero utilizan la aproximación continua a la verosimilitud. Los resultados que se obtienen con las verosimilitudes discretizada y restringida, en este escenario, se muestran en las Tablas 3.16, 3.17 y 3.18.

- **Frecuencia de cobertura de los intervalos de verosimilitud para los parámetros de un modelo Weibull.**

Cuando se usa la verosimilitud discretizada y se estima el parámetro umbral con muestras de tamaño $n \geq 100$, se tienen coberturas de 0.90 a 0.93. Para $n < 100$, las coberturas son apenas aceptables, ya que van de un 0.85 a 0.91. Ahora, cuando se utiliza la verosimilitud restringida los intervalos de verosimilitud perfil del parámetro umbral μ tienen coberturas ligeramente mayores que la discretizada, para $n \leq 50$; sin embargo ocurre lo contrario para $n = 100$, donde la discretizada tiene mejores coberturas. En el caso del umbral, las mejores coberturas se observan cuando el tamaño de muestra aumenta y la precisión corresponde a un valor $2h$ más grande, para ambas verosimilitudes.

Las coberturas de los intervalos de verosimilitud perfil de σ y de β son razonables y buenas, para todos los tamaños de muestras, precisiones considerados y con ambas verosimilitudes. En el caso del parámetro σ las coberturas fluctúan entre 0.93 y aproximadamente 0.95. En cuanto β , su cobertura oscila entre 0.88 y 0.94 con la verosimilitud discretizada y entre 0.89 y 0.95, cuando se usa la verosimilitud restringida.

- **Frecuencia de cobertura de los intervalos de verosimilitud para $Q_{0.05}$.**

Los intervalos de verosimilitud perfil del cuantil $Q_{0.05}$, construidos bajo un modelo Weibull, muestran una cobertura que oscila entre 0.87 y 0.92, aumentando conforme el tamaño de muestra crece y variando muy poco al cambiar la precisión del instrumento. Este comportamiento se observa con ambas verosimilitudes. Estas coberturas son muy parecidas a las que resultan para el parámetro umbral, en este escenario, ya que las muestras simuladas bajo estos valores de parámetros no presentaron problemas en el proceso de optimización.

- **Número de intervalos que sobreestiman o subestiman cierto parámetro.**

Al igual que ocurre en todos los escenarios y con cualquiera de las verosimilitudes utilizadas, la mayoría de los intervalos que no cubren el parámetro umbral están ubicados a la derecha de éste, indicando que $\hat{\mu}$ sobrestima a μ . Con respecto al parámetro $Q_{0.05}$, ocurre el mismo comportamiento. Las frecuencias de cobertura de los intervalos de verosimilitud del umbral y del cuantil $Q_{0.05}$, son muy parecidos, y con ambas verosimilitudes son razonablemente buenos, cuando se tienen muestras de tamaño 50 o mayor.

- **Características de los emv de μ y de β .**

Con la verosimilitud discretizada, para un 33 y 64% de las muestras simuladas ocurrió un emv de μ que cayó en el intervalo que rodea a la observación más chica. Hubo además, hasta un 0.6% de las muestras cuyo $\hat{\mu}$ fue más grande que la observación más pequeña. Estos últimos casos sólo es posible detectarlos con la verosimilitud discretizada. Con la verosimilitud restringida y para muestras de tamaño 100, se llega a tener hasta un 85% de muestras donde se estima $\hat{\mu} = x_{(1)} - h$.

El porcentaje de muestras con $\hat{\beta} > 1$ es muy similar, con ambas verosimilitudes. Cuando $n > 50$, se estima muy bien el parámetro de forma, pues más del 90% de las estimaciones de β fueron mayores o iguales a uno, con cualquiera de las verosimilitudes utilizadas. Este porcentaje disminuye cuando se utilizan muestras de tamaño $n = 25$.

Smith y Weissman (1985) reportan que la frecuencia de muestras problemáticas, con forma de J , aumenta cuando el tamaño de muestra disminuye. Esto puede verse en la columna $\hat{\beta} < 1$, de la Tabla 3.17, para la verosimilitud restringida, donde se observa ese comportamiento. Cabe aclarar que la columna $\hat{\beta} \geq 1$ puede incluir algunos casos de verosimilitudes perfil con esta forma. Para $n = 100$, estos autores encontraron que un 16% de las muestras tenían una verosimilitud perfil L_C del umbral con forma de J , para las cuales no pudieron estimar los parámetros. Cuando disminuyen el tamaño a 25, el porcentaje de estas verosimilitudes perfil, aumenta a un 63% y no reportan, para este escenario, verosimilitudes perfil del umbral de forma aplanada.

- **Frecuencias del cobertura de los intervalos de verosimilitud para el parámetro de forma c de la DGVE.**

Con ambas verosimilitudes, se observa que sólo cuando se consideran muestras de tamaño $n = 25$ y en menos del 1% de los casos podría ser razonable un modelo Gumbel, como indican las frecuencias de muestras con $R_p(c = 0) > 0.1465$ mostradas en la Tabla 3.18. Esto es, en este escenario se identifica claramente el modelo Weibull a partir del cual se simularon los datos.

3.6.5 Resumen general sobre las simulaciones

1. **Velocidad de convergencia de las frecuencias de cobertura conforme n aumenta.** Tanto con la verosimilitud discretizada, como con la verosimilitud restringida, las frecuencias de cobertura de los intervalos de verosimilitud perfil construidos para los parámetros del modelo Weibull convergen a los valores teóricos esperados para modelos regulares, pero con tasa de convergencia diferente. Cuando $\beta < 1$ se necesitan muestras muy grandes (500 o más) para observar valores muy cercanos a las coberturas teóricas de un parámetro umbral, pero conforme β es mayor y distante de uno, la tasa de convergencia es más rápida y las coberturas teóricas pueden observarse desde muestras moderadas e inclusive pequeñas.
2. **Coberturas de los intervalos de verosimilitud de σ y β .** En todos los escenarios, las coberturas para los parámetros σ y β fueron prácticamente las mismas, con ambas verosimilitudes. Las coberturas resultan bastante buenas (mayores que 0.90), inclusive en los casos donde se simulaban muestras con un parámetro de forma $\beta < 1$. Los resultados que asintóticamente se esperan en casos regulares, con la verosimilitud discretizada se dan aún con muestras pequeñas ($n = 25$). Con la verosimilitud restringida se observa un comportamiento similar.
3. **Coberturas de los intervalos de verosimilitud del umbral μ .** Por lo general, para los tamaños de muestras explorados, las coberturas son más bajas para μ que para los otros parámetros, indicando que el umbral es difícil de estimar. Para este parámetro, las coberturas que se obtienen con la verosimilitud discretizada, son muy similares a las que resultan con la verosimilitud restringida y son mayores que 0.90 para valores de $n \geq 100$, si el parámetro de forma $\beta > 1$. Sin embargo, cuando $\beta < 1$, son necesarias muestras de tamaño 500 o más, para tener valores aproximados a los teóricos. En los Escenarios 1a y 1b, las coberturas de la verosimilitud restringida son un poco mejores que las obtenidas con la verosimilitud discretizada. La razón de esto es que los intervalos generados con la restringida son un poco más anchos que los que se obtienen con la verosimilitud discretizada.
4. **Verosimilitudes perfil con forma de J .** Solamente la verosimilitud perfil restringida presenta una forma de J , aunque sin singularidad alguna, ya que el parámetro μ siempre toma valores menores o iguales que $x_{(1)} - h$. Estas formas ocurren más frecuentemente en modelos donde β es menor o cercano a 1, se tienen muestras pequeñas y/o la estimación del umbral resulta $\hat{\mu} = x_{(1)} - h$. Las muestras donde la restringida arroja las estimaciones $\hat{\beta} < 1$, serán por lo general muestras problemáticas cuando se analicen con la verosimilitud continua L_C , ya que ésta no es acotada y el máximo global se alcanza en la singularidad. Este problema se evita al utilizar la verosimilitud discretizada o la restringida.

Entre el 99 y 100% de las muestras simuladas en el Escenario 1 y analizadas con

la verosimilitud restringida, tuvieron una estimación $\hat{\beta} < 1$. En el Escenario 4 donde se satisface que $1 < \beta < 2$, se observa hasta un 33% de muestras donde $\hat{\beta} < 1$ y se presentan más cuando el tamaño de muestra es pequeño y $2h$ disminuye. En el Escenario 2, donde $\beta = 2$, ocurren pocas de estas muestras, aproximadamente un 4% cuando $n = 25$ y ninguna para $n = 100$. Finalmente, en el Escenario 3, el porcentaje de estas muestras fue menor del 0.04% y sólo ocurrieron para $n = 25$.

5. **Estimación puntual del parámetro umbral.** Con la verosimilitud discretizada, en particular para los Escenarios 1a, 1b y 4, hay proporciones importantes de muestras donde el emv $\hat{\mu}$ cae dentro del intervalo que rodea a la observación más pequeña $x_{(1)}$. Esta situación es única y distingue a la verosimilitud discretizada de las demás, pues incluso puede ocurrir que $\hat{\mu}$ sea mayor que $x_{(1)}$. Las muestras donde la verosimilitud discretizada reporta un emv de μ en el intervalo $(x_{(1)}, x_{(1)} + h)$, son muestras donde la aproximación continua L_C resulta no-acotada. En los Escenarios 2 y 3 casi no ocurre esta situación pues la estimación del parámetro de forma es generalmente mayor de 1. Estas estimaciones del umbral no pueden obtenerse con la verosimilitud restringida; en esos casos, la verosimilitud restringida generalmente indica que $\hat{\mu} = x_{(1)} - h$. Por otra parte, para el Escenario 3, el número de muestras donde $\hat{\mu}$ cae a la izquierda de $x_{(1)} - h$, es muy parecido con ambas verosimilitudes.

6. **Coberturas de los intervalos de verosimilitud del $Q_{0.05}$.** Las coberturas de los intervalos de verosimilitud perfil del cuantil $Q_{0.05}$ de un modelo Weibull, son prácticamente las mismas con ambas verosimilitudes. Sólo en el Escenario 1 son ligeramente mejores cuando se utiliza la verosimilitud discretizada. En las muestras simuladas se observa que para garantizar una cobertura del 90% de los intervalos perfiles de este cuantil, debe ocurrir lo siguiente: si $\beta < 1$ se requiere que $n \geq 100$; si $1 \leq \beta < 2$, puede darse desde $n \geq 50$, pero depende de la precisión del instrumento de medición; finalmente, si $\beta > 2$, basta con $n \geq 25$. Con ambas verosimilitudes se observa que los intervalos de verosimilitud perfil que no cubren a $Q_{0.05}$, lo sobreestiman, como ocurre con el parámetro umbral. Así pues, en escenarios donde se tenga una estimación $\hat{\beta} < 1$, se recomienda contar con muestras de tamaño $n \geq 100$, o bien estimar un cuantil de probabilidad mayor.

7. **Coberturas de los intervalos de verosimilitud de c .** Todos los escenarios están asociados a modelos Weibull que son claramente distintos de uno Gumbel, excepto por el Escenario 3, donde el modelo Weibull utilizado está muy cercano a su modelo límite Gumbel. Las coberturas empíricas de los intervalos de verosimilitud perfil de c que excluyeron a $c = 0$, con ambas verosimilitudes, son altas; del 87 al 100% en los Escenarios 1a, 1b, 2 y 4. Esto se revierte en el Escenario 3, pues casi todos los intervalos de verosimilitud discretizada y todos los de la restringida incluyen a $c = 0$, lo cual indica que un modelo Gumbel también es razonable para describir a los datos. En este escenario, la verosimilitud discretizada identifica entre un 6 y 11% de las muestras como modelos

Weibull y no como modelos Gumbel; los porcentajes más altos ocurren cuando $n = 100$.

8. **Efecto de la precisión en las frecuencias de cobertura.** En todos los escenarios y para ambas verosimilitudes, se observó que al fijar la precisión $2h$, las frecuencias de cobertura calculadas aumentan conforme crece el tamaño de la muestra. En el caso de la verosimilitud discretizada, las coberturas convergen a las predichas bajo la distribución teórica correspondiente que se obtiene en los casos regulares. Un comportamiento similar se observa con la verosimilitud restringida.

Cuando se fija un tamaño de muestra y se varía la precisión del instrumento, el comportamiento de las coberturas depende de cómo son los valores de los parámetros del modelo simulado. Por ejemplo, en los escenarios donde los datos provienen de una distribución donde $\beta < 1$ o $\beta = 1.25$, las coberturas resultaron mejores, para parámetros como el umbral o un cuantil, conforme el valor $2h$ aumenta. En cambio, se mantienen coberturas muy similares, dentro de un mismo tamaño de muestra, para parámetros como σ y β . Cuando $\beta \geq 2$ la precisión del instrumento no parece afectar mucho las coberturas obtenidas.

Tomando en cuenta lo expuesto anteriormente, se concluye que tanto la verosimilitud discretizada como la restringida, permiten dar buenas inferencias, con muestras de tamaño $n \geq 25$, para los parámetros σ y β de una distribución Weibull de tres parámetros y para cuantiles de probabilidad mayor a 0.05. El parámetro umbral es difícil de estimar y se requieren muestras con $n \geq 500$, para poder obtener inferencias razonables, tanto con la verosimilitud discretizada como con la restringida.

Un hecho importante es que el comportamiento asintótico de la razón de verosimilitud discretizada, en una distribución con umbral y singularidad, como la distribución Weibull de tres parámetros, resulta similar al obtenido con un modelo regular, como el multinomial. Se reitera que para la verosimilitud perfil del umbral, la velocidad de convergencia a los resultados asintóticos es mucho más lenta que para otros parámetros del modelo; por ello, se requieren muestras más grandes para obtener inferencias razonables.

Aunque ambas verosimilitudes se comportan de manera muy similar, en cuanto a la estimación de los parámetros Weibull μ , σ , β y un cuantil Q_α , en todos los escenarios analizados y ambas resuelven el problema de estimación que L_C presenta cuando se utilizan densidades con un parámetro umbral y una singularidad, la verosimilitud discretizada posee las siguientes ventajas:

1. La verosimilitud discretizada siempre es acotada, pues está definida en términos de probabilidades y nunca tiene forma de J .
2. La verosimilitud discretizada permite estimar de manera más precisa el parámetro umbral μ , pues con ella es posible obtener estimaciones mayores que la observación más pequeña registrada.

3. Los intervalos de verosimilitud discretizada perfil suelen ser un poco más angostos que los de la restringida, si bien tienen aproximadamente las mismas coberturas razonables. Esto se da en particular para modelos Weibull donde $\beta < 1$.
4. Aún con muestras pequeñas, la verosimilitud discretizada permite discriminar mejor entre modelos distintos (como el Weibull y su modelo Gumbel límite), como se observó en el Escenario 3.
5. La verosimilitud discretizada incorpora, de una manera sencilla, la precisión del instrumento de medición en cada una de sus observaciones. Esto es importante, pues en la práctica pueden presentarse situaciones como la descrita en el Ejemplo 4.1.3, donde cada una de las observaciones puede tener una precisión diferente.

Gracias al uso de la estadística computacional, se pudo implementar la verosimilitud discretizada propuesta, para los diferentes escenarios seleccionados. El análisis obtenido permitió identificar qué tamaños de muestras son necesarios para que las estimaciones de los parámetros del modelo Weibull se puedan considerar razonables. Si bien las verosimilitudes discretizada y restringida arrojan inferencias similares para los parámetros de los modelos analizados, se recomienda utilizar la verosimilitud discretizada propuesta en esta tesis, a la luz de las ventajas antes mencionadas y por la facilidad de cómputo que hoy en día se tiene.

Escenario 1a		n	$2h$	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$FC(\sigma)$	# de I.V. a la izq. de σ	# de I.V. a la derecha de σ	$FC(\beta)$	# de I.V. a la izq. de β	# de I.V. a la derecha de β
$\mu = 4$	0.01	6285	0	3715	9568	372	60	9590	378	32	32	32
	0.005	5488	0	4512	9612	333	55	9568	410	22	22	22
	0.001	3778	0	6222	9621	319	60	9551	420	29	29	29
$\sigma = 18$	0.01	7120	0	2880	9575	302	123	9578	308	114	114	114
	0.005	6492	0	3508	9566	308	126	9608	307	85	85	85
	0.001	4861	0	5139	9533	333	134	9574	338	88	88	88
$\beta = 0.7$	0.01	7897	0	2103	9517	285	198	9560	187	253	253	253
	0.005	7259	0	2741	9490	315	195	9553	293	154	154	154
	0.001	5890	0	4110	9504	330	166	9559	330	111	111	111
500	0.01	9335	0	665	9509	268	223	9522	262	216	216	216

Escenario 1a		n	$2h$	$FC(Q_{0.05})$	# de I.V. a la izq. de $Q_{0.05}$	# de I.V. a la derecha de $Q_{0.05}$	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$	# de I.V. a la izq. de $x_{(1)} - h$	# de I.V. a la derecha de $x_{(1)} + h$
$\mu = 4$	0.01	7951	3	2046	49	9951	49	9931	20	20
	0.005	7707	3	2290	38	9962	38	9957	5	5
	0.001	7290	0	2710	28	9972	28	9972	0	0
$\sigma = 18$	0.01	8762	8	1230	6	9994	4	9966	30	30
	0.005	8774	4	1222	11	9989	10	9982	8	8
	0.001	8749	1	1250	1	9999	0	9999	0	0
$\beta = 0.7$	0.01	9495	18	487	0	10000	0	9875	125	125
	0.005	9566	5	429	0	10000	0	9979	21	21
	0.001	9603	2	395	0	10000	0	10000	0	0
500	0.01	9847	7	146	0	10000	0	7601	2399	2399

Escenario 1a		n	$2h$	$FC(Q_{0.10})$	# de I.V. a la izq. de $Q_{0.10}$	# de I.V. a la derecha de $Q_{0.10}$	$FC(Q_{0.50})$	# de I.V. a la izq. de $Q_{0.50}$	# de I.V. a la derecha de $Q_{0.50}$
$\mu = 4$	0.01	8448	289	1263	9527	384	89	89	89
	0.005	8416	329	1255	9517	405	78	78	78
	0.001	8330	446	1224	9514	442	44	44	44
$\sigma = 18$	0.01	9249	264	487	9573	302	125	125	125
	0.005	9295	246	459	9545	319	136	136	136
	0.001	9283	349	368	9520	385	95	95	95
$\beta = 0.7$	0.01	9452	236	312	9506	277	217	217	217
	0.005	9412	306	282	9488	319	193	193	193
	0.001	9456	311	233	9491	364	145	145	145
500	0.01	9490	263	247	9555	256	209	209	209

Tabla 3.5: Escenario 1a bajo la verosimilitud discretizada.

Escenario 1a	n	$2h$	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$\hat{\mu} < x(1) - h$	$\hat{\mu} = x(1) - h$	$FC(\sigma)$	$FC(\beta)$
$\mu = 4$	25	0.01	6994	0	3006	55	9945	9578	9677
		0.005	6298	0	3702	39	9961	9671	9732
		0.001	4728	0	5272	38	9962	9684	9552
$\sigma = 18$	50	0.01	7643	0	2357	6	9994	9588	9623
		0.005	7069	0	2931	11	9989	9639	9731
		0.001	5709	0	4291	1	9999	9587	9616
$\beta = 0.7$	100	0.01	8271	0	1729	1	9999	9527	9595
		0.005	7687	0	2313	1	9999	9635	9713
		0.001	6505	0	3495	0	10000	9535	9605
	500	0.01	9475	0	525	0	10000	9513	9555

Escenario 1a	n	$2h$	$FC(Q_{0.05})$	# de I.V. a la izq. de $Q_{0.05}$	# de I.V. a la derecha de $Q_{0.05}$	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$
$\mu = 4$	25	0.01	7548	107	2345	50	9950
		0.005	7858	95	2047	33	9967
		0.001	6955	133	2912	37	9963
$\sigma = 18$	50	0.01	8402	91	1507	7	9993
		0.005	8425	72	1503	9	9991
		0.001	8458	128	1414	1	9999
$\beta = 0.7$	100	0.01	9273	56	671	0	10000
		0.005	9342	58	600	0	10000
		0.001	9374	70	556	0	10000
	500	0.01	9614	0	386	0	10000

Tabla 3.6: Escenario 1a bajo la verosimilitud restringida.

Escenario 1b		n	$2h$	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$FC(\sigma)$	$FC(\beta)$	$\hat{\mu} < x_{(1)} - h$	$\hat{\mu} \in [x_{(1)} - h, x_{(1)})$	$\hat{\mu} \in [x_{(1)}, x_{(1)} + h]$
$\mu = 4.0073$ $\sigma = 18$ $\beta = 0.7$	25	0.01	6312	0	3688	9582	9634	103	9876	21	
	50	0.01	7119	6	2875	9529	9617	60	9904	36	
	100	0.01	7875	13	2112	9548	9602	16	9838	146	
	500	0.01	9276	144	580	9480	9485	2	4906	5092	

Escenario 1b		n	$2h$	$FC(Q_{0.05})$	# de I.V. a la izq. de $Q_{0.05}$	# de I.V. a la derecha de $Q_{0.05}$	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$
$\mu = 4.0073$ $\sigma = 18$ $\beta = 0.7$	25	0.01	7992	4	2004	33	0	
	50	0.01	8800	9	1191	0	10000	
	100	0.01	9543	10	447	0	10000	
	500	0.01	9881	5	114	0	10000	

Tabla 3.7: Escenario 1b bajo la verosimilitud discretizada, con $\mu = 4.0073$.

Escenario 1b		n	$2h$	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$FC(\sigma)$	$FC(\beta)$	$\hat{\mu} < x_{(1)} - h$	$\hat{\mu} = x_{(1)} - h$
$\mu = 4.0073$ $\sigma = 18$ $\beta = 0.7$	25	0.01	7014	0	2986	9623	9751	43	9957	
	50	0.01	7620	0	2380	9581	9701	12	9988	
	100	0.01	8281	0	1719	9660	9724	2	9998	
	500	0.01	9639	0	361	9827	9801	0	10000	

Escenario 1b		n	$2h$	$FC(Q_{0.05})$	# de I.V. a la izq. de $Q_{0.05}$	# de I.V. a la derecha de $Q_{0.05}$	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$
$\mu = 4.0073$ $\sigma = 18$ $\beta = 0.7$	25	0.01	7616	88	2296	33	9967	
	50	0.01	8449	86	1465	0	10000	
	100	0.01	9306	73	621	0	10000	
	500	0.01	9614	0	386	0	10000	

Tabla 3.8: Escenario 1b bajo la verosimilitud restringida, con $\mu = 4.0073$.

Escenario 1a	n	$2h$	Verosimilitud Discretizada			Verosimilitud Restringida		
			# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$	# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$
$\mu = 4$	25	0.01	1	9999	0	0	10000	0
		0.005	1	9999	0	0	10000	0
		0.001	3	9997	0	0	10000	0
	50	0.01	0	10000	0	0	10000	0
		0.005	0	10000	0	0	10000	0
		0.001	0	10000	0	0	10000	0
100	0.01	0	10000	0	0	10000	0	
	0.005	0	10000	0	0	10000	0	
	0.001	0	10000	0	0	10000	0	

Tabla 3.9: Escenario 1: Análisis utilizando la DGVE.

Escenario 2	n	2h	FC(μ)		# de I.V. a la derecha de μ		FC(σ)		# de I.V. a la izquierda de σ		FC(β)		# de I.V. a la izquierda de β		# de I.V. a la derecha de β	
			izq.	deq.	izq.	deq.	izq.	deq.	izq.	deq.	izq.	deq.	izq.	deq.		
$\mu = 8$	25	0.1	9138	51	811	9172	762	66	9119	733	148					
		0.05	9151	58	791	9121	796	83	9144	720	736					
		0.01	9156	67	777	9138	775	87	9124	720	157					
$\sigma = 6$	50	0.1	9344	51	605	9307	609	84	9320	533	147					
		0.05	9342	50	608	9291	636	73	9334	540	126					
		0.01	9320	45	635	9293	630	77	9341	545	114					
$\beta = 2$	100	0.1	9412	41	547	9374	547	79	9407	469	124					
		0.05	9436	55	509	9393	512	95	9365	490	145					
		0.01	9363	71	566	9364	538	98	9377	476	147					

Escenario 2	n	2h	FC($Q_{0.05}$)		# de I.V. a la derecha de $Q_{0.05}$		$\hat{\beta} \geq 1$	$\hat{\beta} < 1$	$\hat{\mu} < x_{(1)} - h$	$\hat{\mu} \in [x_{(1)} - h, x_{(1)})$	$\hat{\mu} \in [x_{(1)}, x_{(1)} + h]$
			izq.	deq.	izq.	deq.					
$\mu = 8$	25	0.1	9070	155	775	9928	72	9602	398	0	
		0.05	9112	120	768	9867	133	9648	352	0	
		0.01	9126	138	736	9631	369	9554	446	0	
$\sigma = 6$	50	0.1	9307	131	562	10000	0	9977	23	0	
		0.05	9277	155	568	9999	1	9986	14	0	
		0.01	9267	136	597	9997	3	9990	10	0	
$\beta = 2$	100	0.1	9332	127	541	10000	0	9998	2	0	
		0.05	9395	147	458	10000	0	10000	0	0	
		0.01	9355	148	497	10000	0	10000	0	0	

Tabla 3.10: Escenario 2 bajo la verosimilitud discretizada.

Escenario 2		n	$2h$	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$\hat{\mu} < x_{(1)} - h$	$\hat{\mu} = x_{(1)} - h$	$FC(\sigma)$	$FC(\beta)$	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$
$\mu = 8$	0.1	100	0.1	9076	100	824	9100	900	9097	9050	9949	51
	0.05	62	0.05	9145	62	793	9432	568	9115	9131	9890	110
	0.01	67	0.01	9157	67	776	9529	471	9137	9122	9673	327
$\sigma = 6$	0.1	184	0.1	9201	184	615	9868	132	9152	9162	10000	0
	0.05	66	0.05	9325	66	609	9952	48	9257	9303	9999	1
	0.01	46	0.01	9319	46	635	9988	12	9292	9339	9997	3
$\beta = 2$	0.1	271	0.1	9179	271	550	9992	8	9129	9156	10000	0
	0.05	105	0.05	9386	105	509	9998	2	9323	9294	10000	0
	0.01	71	0.01	9363	71	566	10000	0	9361	9375	10000	0

Escenario 2		n	$2h$	$FC(Q_{0.05})$	# de I.V. a la izq. de $Q_{0.05}$	# de I.V. a la derecha de $Q_{0.05}$
$\mu = 8$	0.1	25	0.1	9058	111	831
	0.05	25	0.05	9109	96	795
	0.01	25	0.01	9127	111	762
$\sigma = 6$	0.1	50	0.1	9303	16	681
	0.05	50	0.05	9276	15	709
	0.01	50	0.01	9266	16	718
$\beta = 2$	0.1	100	0.1	9330	0	670
	0.05	100	0.05	9393	1	606
	0.01	100	0.01	9355	0	645

Tabla 3.11: Escenario 2 bajo la verosimilitud restringida.

Escenario 2	n	$2h$	Verosimilitud Discretizada			Verosimilitud Restringida		
			# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$	# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$
$\mu = 8$	25	0.1	1265	8735	0	1263	8737	0
		0.05	1293	8707	0	1291	8709	0
		0.01	1302	8698	0	1302	8698	0
$\sigma = 6$	50	0.1	56	9944	0	56	9944	0
		0.05	50	9950	0	50	9950	0
		0.01	47	9953	0	47	9953	0
$\beta = 2$	100	0.1	0	10000	0	0	10000	0
		0.05	0	10000	0	0	10000	0
		0.01	0	10000	0	0	10000	0

Tabla 3.12: Escenario 2: Análisis utilizando la DGVE.

Escenario 3		n	$2h$	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$FC(\sigma)$	# de I.V. a la izq. de σ	# de I.V. a la derecha de σ	$FC(\beta)$	# de I.V. a la izq. de β	# de I.V. a la derecha de β
$\mu = -3.5$	0.01 0.005 0.001	25	0.01	9598	2	400	9599	399	2	9600	398	2
			0.005	9557	2	441	9547	441	2	9566	433	1
			0.001	9585	0	415	9585	0	0	9592	408	0
$\sigma = 5$	0.01 0.005 0.001	50	0.01	9653	6	341	9651	343	6	9651	343	6
			0.005	9667	7	326	9666	326	8	9681	310	9
			0.001	9682	7	311	9681	312	7	9693	299	8
$\beta = 20$	0.01 0.005 0.001	100	0.01	9654	32	314	9654	314	32	9657	312	31
			0.005	9670	27	303	9670	303	27	9678	298	24
			0.001	9682	27	291	9681	291	28	9684	287	29

Escenario 3		n	$2h$	$FC(Q_{0.05})$ c/modelo Weibull	# de I.V. izq. de $Q_{0.05}$ (Wbl)	# de I.V. derecha de $Q_{0.05}$ (Wbl)	$FC(Q_{0.05})$ c/modelo Gumbel	# de I.V. derecha de $Q_{0.05}$ (Gum)	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$	$\hat{\mu} < x_{(1)} - h$	$\hat{\mu} \in [x_{(1)} - h, x_{(1)})$	$\hat{\mu} \in [x_{(1)}, x_{(1)} + h]$
$\mu = -3.5$	0.01 0.005 0.001	25	0.01	9018	224	758	9488	512	10000	0	9991	9	0
			0.005	8908	215	876	9408	592	9999	1	9991	9	0
			0.001	8975	218	807	9495	505	9996	4	9994	6	0
$\sigma = 5$	0.01 0.005 0.001	50	0.01	9217	182	601	9480	520	10000	0	10000	0	0
			0.005	9199	200	601	9482	518	10000	0	10000	0	0
			0.001	9224	210	566	9472	528	10000	0	10000	0	0
$\beta = 20$	0.01 0.005 0.001	100	0.01	9407	196	397	9504	496	10000	0	10000	0	0
			0.005	9383	183	434	9450	550	10000	0	10000	0	0
			0.001	9403	201	396	9481	519	10000	0	10000	0	0

Tabla 3.13: Escenario 3 bajo la verosimilitud discretizada.

La función de verosimilitud discretizada

Escenario 3											
$\mu = -3.5$	$\sigma = 5$	$\beta = 20$	n	$2h$	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$\hat{\mu} < x_{(1)} - h$	$\hat{\mu} = x_{(1)} - h$	$FC(\sigma)$	$FC(\beta)$
0.01 0.005 0.001	25	0.01 0.005 0.001	2	399	9599	1	441	9987	13	9599	9601
			1	441	9558	0	415	9990	10	9558	9565
			0	415	9585	0	341	9993	7	9585	9592
0.01 0.005 0.001	50	0.01 0.005 0.001	6	341	9653	7	326	10000	0	9651	9651
			7	326	9667	0	311	10000	0	9665	9681
			0	311	9682	0	315	10000	0	9681	9693
0.01 0.005 0.001	100	0.01 0.005 0.001	34	315	9651	27	303	10000	0	9653	9657
			27	303	9670	0	291	10000	0	9670	9678
			28	291	9681	0	288	10000	0	9680	9684

Escenario 3											
$\mu = -3.5$	$\sigma = 5$	$\beta = 20$	n	$2h$	$FC(Q_{0.05})$ c/modelo Weibull	# de I.V. a la izq. de $Q_{0.05}$ (Wbl)	# de I.V. a la derecha de $Q_{0.05}$ (Wbl)	$FC(Q_{0.05})$ c/modelo Gumbel	# de I.V. a la derecha de $Q_{0.05}$ (Gum)	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$
0.01 0.005 0.001	25	0.01 0.005 0.001	1	843	9247	0	752	9548	452	10000	0
			0	843	9157	0	805	9531	469	9999	1
			0	805	9195	0	640	9546	454	9996	4
0.01 0.005 0.001	50	0.01 0.005 0.001	0	640	9360	0	640	9496	504	10000	0
			0	640	9360	0	625	9516	484	10000	0
			0	625	9375	0	536	9464	536	10000	0
0.01 0.005 0.001	100	0.01 0.005 0.001	0	516	9484	0	548	9361	639	10000	0
			0	548	9452	0	513	9349	651	10000	0
			0	513	9487	0	668	9332	668	10000	0

Tabla 3.14: Escenario 3 bajo la verosimilitud restringida.

Escenario 3	n	$2h$	Verosimilitud Discretizada			Verosimilitud Restringida		
			# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$	# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$
$\mu = -3.5$	25	0.01	9279	624	97	10000	0	0
		0.005	9200	699	101	10000	0	0
		0.001	9233	669	98	10000	0	0
$\sigma = 5$	50	0.01	9167	782	51	10000	0	0
		0.005	9187	758	55	10000	0	0
		0.001	9214	730	56	10000	0	0
$\beta = 20$	100	0.01	8923	1053	24	10000	0	0
		0.005	8931	1048	21	10000	0	0
		0.001	8939	1026	35	10000	0	0

Tabla 3.15: Escenario 3: Análisis utilizando la DGVE.

Escenario 4		n	$2h$	$FC(\mu)$	# de I.V. a la izq. de μ	# de I.V. a la derecha de μ	$FC(\sigma)$	# de I.V. a la izq. de σ	# de I.V. a la derecha de σ	$FC(\beta)$	# de I.V. a la izq. de β	# de I.V. a la derecha de β
$\mu = 0$	0.02	25	0.02	8817	10	1173	9440	473	87	9177	695	128
			0.01	8660	4	1336	9402	536	62	9021	877	102
			0.005	8546	13	1441	9329	594	77	8837	1054	109
$\sigma = 1$	0.02	50	0.02	9118	0	882	9434	475	91	9363	513	124
			0.01	8964	1	1035	9408	520	72	9215	682	103
			0.005	8856	1	1143	9342	574	84	9150	756	94
$\beta = 1.25$	0.02	100	0.02	9288	0	712	9471	429	100	9473	404	123
			0.01	9157	0	843	9430	451	119	9391	487	122
			0.005	9054	0	946	9327	547	126	9265	618	117

Escenario 4		n	$2h$	$FC(Q_{0.05})$	# de I.V. a la izq. de $Q_{0.05}$	# de I.V. a la derecha de $Q_{0.05}$	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$	$\hat{\mu} < x_{(1)} - h$	$\hat{\mu} \in [x_{(1)} - h, x_{(1)})$	$\hat{\mu} \in [x_{(1)}, x_{(1)} + h]$
$\mu = 0$	0.02	25	0.02	8789	155	1056	7846	2154	4605	5357	38
			0.01	8701	164	1135	7192	2808	5021	4975	4
			0.005	8673	195	1132	6469	3531	5133	4866	1
$\sigma = 1$	0.02	50	0.02	9086	150	764	9470	530	4572	5406	22
			0.01	8978	184	838	9245	755	5748	4250	2
			0.005	8904	203	893	9019	981	6468	3532	0
$\beta = 1.25$	0.02	100	0.02	9194	149	657	9942	58	3088	6851	61
			0.01	9130	179	691	9939	61	5112	4888	0
			0.005	9108	199	693	9899	101	6652	3348	0

Tabla 3.16: Escenario 4 bajo la verosimilitud discretizada.

Escenario 4		n	$2h$	$FC(\mu)$	# of L.I. a la left of μ	# of L.I. a la right of μ	$\hat{\mu} < x_{(1)} - h$	$\hat{\mu} = x_{(1)} - h$	$FC(\sigma)$	$FC(\beta)$	$\hat{\beta} \geq 1$	$\hat{\beta} < 1$
$\mu = 0$	0.02	25	0.02	8862	10	1128	3523	6477	9440	9249	8061	1939
	0.01	25	0.01	8722	3	1275	4400	5600	9404	9073	7439	2561
	0.005	25	0.005	8607	11	1382	4877	5123	9339	8917	6721	3279
$\sigma = 1$	0.02	50	0.02	9123	0	877	2979	7021	9441	9402	9584	416
	0.01	50	0.01	8972	1	1027	4746	5254	9410	9235	9316	684
	0.005	50	0.005	8859	1	1140	5972	4028	9341	9170	9108	892
$\beta = 1.25$	0.02	100	0.02	9273	0	727	1540	8461	9482	9504	9973	27
	0.01	100	0.01	9129	0	871	3698	6302	9423	9398	9952	48
	0.005	100	0.005	9038	0	962	5846	4154	9331	9257	9912	88

Escenario 4		n	$2h$	# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$	$FC(Q_{0.05})$	# de I.V. a la izq. de $Q_{0.05}$	# de I.V. a la derecha de $Q_{0.05}$
$\mu = 0$	0.02	25	0.02	98	9902	0	8771	145	1084
	0.01	25	0.01	84	9916	0	8713	157	1130
	0.005	25	0.005	74	9926	0	8681	190	1129
$\sigma = 1$	0.02	50	0.02	0	10000	0	9064	137	799
	0.01	50	0.01	1	9999	0	8971	163	866
	0.005	50	0.005	0	10000	0	8905	185	910
$\beta = 1.25$	0.02	100	0.02	0	10000	0	9128	12	860
	0.01	100	0.01	0	10000	0	9127	17	856
	0.005	100	0.005	0	10000	0	9115	28	857

Tabla 3.17: Escenario 4 bajo la verosimilitud restringida.

Escenario 4	n	$2h$	Verosimilitud Discretizada			Verosimilitud Restringida		
			# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$	# de I.V. donde $R_p(c=0) > 0.1465$	# de I.V. a la izq. de $c=0$	# de I.V. a la derecha de $c=0$
$\mu = 0$	25	0.02	98	9902	0	98	9916	0
		0.01	84	9916	0	84	9906	0
		0.005	75	9925	0	74	9926	0
$\sigma = 1$	50	0.02	0	10000	0	0	10000	0
		0.01	1	9999	0	1	9999	0
		0.005	0	10000	0	0	10000	0
$\beta = 1.25$	100	0.02	0	10000	0	0	10000	0
		0.01	0	10000	0	0	10000	0
		0.005	0	10000	0	0	10000	0

Tabla 3.18: Escenario 4: Análisis utilizando la DGVE.

Capítulo 4

Aplicaciones de las verosimilitudes restringida y discretizada

En este capítulo se muestra, a través de siete ejemplos prácticos, cómo la verosimilitud perfil es una herramienta fundamental en la inferencia científica. En todos estos ejemplos se consideran densidades con un parámetro umbral y una singularidad, por lo que se trata de modelos no regulares. El enfoque estadístico tradicional, que utiliza la aproximación continua a la función de verosimilitud L_C , origina verosimilitudes no acotadas y por tanto la estimación es problemática. Esto ha dado lugar a críticas injustificadas de la función de verosimilitud. Como se muestra en estos ejemplos, el problema de verosimilitudes no acotadas se resuelve al utilizar la verosimilitud discretizada, propuesta en esta tesis, o bien la verosimilitud restringida.

Otro tipo de críticas se ha dado para el caso de que la verosimilitud perfil es aplanada, la cual se ha calificado incorrectamente como no informativa. En los ejemplos que se incluyen en este capítulo, se muestra que esta forma indica que podría haber modelos más sencillos que describan bien a los datos.

En todos los ejemplos se mostrará cómo las verosimilitudes perfil restringida y discretizada, calculadas para un parámetro umbral, un cuantil adecuado, un parámetro de forma, o parámetros resultantes de una transformación, como $\tau = P(Y > c)$, son de gran utilidad y permiten dar respuesta a las preguntas que son de interés. En cada uno de los ejemplos se efectúa una comparación entre la verosimilitud restringida y la discretizada y se observa que, independientemente del parámetro a estimar, ambas verosimilitudes producen inferencias razonables para casi todos los parámetros, incluso en muestras pequeñas. En el caso del umbral se puede requerir, según sea el parámetro de forma de la densidad, el contar con muestras más grandes para lograr una buena estimación; por ello se sugiere considerar también algún cuantil pequeño cuya estimación sea razonable con el tamaño de muestra considerado.

Los ejemplos se presentan de acuerdo a las tres formas típicas que toma la verosimilitud perfil del parámetro umbral cuando se usa la aproximación continua L_C , mismas que se mostraron en la Figura 2.1.

- **Ejemplos Tipo I.** En estos ejemplos la verosimilitud perfil del umbral tiene la forma presentada en Figura 2.1(a) y describen situaciones donde L_C tiene un

máximo local y posee una singularidad que hereda de la función de densidad. Cuando se utiliza la verosimilitud restringida o bien la discretizada, la forma de la verosimilitud perfil del umbral es siempre acotada y generalmente unimodal y acampanada, aunque puede ser asimétrica. El emv que se obtiene con estas verosimilitudes coincide con el punto donde se alcanza un máximo local con L_C . Para ilustrar este tipo de situación se presentan tres ejemplos. El primero es un conjunto de datos donde se estudian los diámetros de árboles de cerezas negras. El segundo considera los tiempos de espera hasta poder detectar cáncer en ratas a las que se les inyectó una sustancia cancerígena. En el tercer ejemplo se estudian datos del índice de masa corporal de mujeres sonorenses.

- **Ejemplo Tipo II.** Su verosimilitud perfil corresponde a la mostrada en la Figura 2.1(b). Estos son casos donde la verosimilitud perfil del umbral obtenida con L_C tiene forma de J y no está acotada. La función de densidad conjunta tiene una singularidad que le hereda a L_C . Estas son situaciones en que se ha criticado injustificadamente a la función de verosimilitud, argumentando que no existe el estimador máximo verosímil de los parámetros. Al utilizar funciones de verosimilitud adecuadas que sí están acotadas, como L_r o L_D es posible maximizarlas sin dificultad alguna. L_r sigue teniendo forma de J , si bien está acotada. Por otra parte, L_D si tiene un máximo global. El ejemplo que se presenta considera los tiempos de espera que ratas de laboratorio tardan para realizar una tarea de aprendizaje. A pesar de que la forma de J se menciona en varios artículos, en la literatura revisada no se encontró algún ejemplo real con esta forma de verosimilitud perfil umbral, por lo que el ejemplo que aquí se presenta es de interés particular.

- **Ejemplos Tipo III.** En estos ejemplos, la verosimilitud perfil del umbral corresponde a la mostrada en la Figura 2.1(c). Esta forma describe situaciones donde la aproximación continua L_C tiene un máximo local alejado de la singularidad y la verosimilitud perfil del parámetro umbral es muy aplanada lejos de la singularidad y muy asimétrica. Injustificadamente se le ha calificado como no informativa. Se ha dicho también que los emv son inestables, por ser difíciles de encontrar numéricamente, debido a la forma aplanada de la verosimilitud. En realidad, esta forma indica que existe un modelo límite más sencillo que describe bien a los datos. Para mostrar esta situación se presentan tres ejemplos. El primero consiste de los máximos anuales del nivel del mar en Port Pirie, Australia. El segundo es un conjunto de datos de los voltajes de rompimiento de un fluido para evaluar la resistencia que éste tenía como aislante frente al paso de corriente eléctrica. Finalmente, en el tercer ejemplo se analiza la dureza de fibras de vidrio.

4.1 Ejemplos tipo I

4.1.1 Diámetros de árboles de cerezas negras

En este problema se analizan los diámetros (en pulgadas) de 31 árboles adultos de cerezas negras (Ryan et al., 1985), medidos 54 pulgadas arriba del nivel del suelo. La observación mínima es $x_{(1)} = 8.3$ y la máxima $x_{(31)} = 20.6$; la media y varianzas muestrales son $\bar{x} = 13.25$ y $s^2 = 9.85$, respectivamente. Los diámetros se presentan en la Sección C.1 del Apéndice C y están reportados con un decimal, por lo que la precisión del instrumento de medición se consideró como $2h = 0.1$. De hecho, la distancia mínima entre dos observaciones consecutivas, no repetidas, es de 0.1, lo cual confirma el valor de h seleccionado.

En ciencias forestales es usual utilizar un modelo Weibull (Bailey and Dell, 1973) para este tipo de datos. El parámetro umbral μ es muy informativo como cota inferior de estos diámetros y en general es un parámetro de interés. Otro parámetro que proporciona información similar es un cuantil pequeño. Se tiene la ventaja de que el modelo Weibull se puede reparametrizar fácilmente en términos de un cuantil; además, generalmente se simetriza la función de verosimilitud y se logra así que la estimación sea mucho más eficiente. Para estos datos, la aproximación continua a la verosimilitud L_C tiene un máximo local y la singularidad se alcanza en $\mu = x_{(1)}$.

Como se podrá ver en las figuras presentadas, la verosimilitud perfil relativa restringida R_r , que se muestra con una línea punteada, y la verosimilitud relativa discretizada R_D , dibujada con una línea continua, son prácticamente iguales para todos los parámetros analizados, por lo que se traslapan en su mayoría. Esto se observa tanto cuando se utiliza un modelo Weibull o cuando se utiliza la DGVE. De hecho, los máximos globales de R_r y R_D , son prácticamente iguales al máximo local que se alcanza con la aproximación continua, por lo que las estimaciones máximo verosímiles de estos parámetros, coinciden en las tres verosimilitudes.

La Figura 4.1 muestra las verosimilitudes perfiles relativas, restringida y discretizada, para los parámetros μ , σ y β del modelo Weibull estimado, donde puede verse que éstas se traslapan por completo. Estas tres verosimilitudes son unimodales y ligeramente asimétricas.

Los intervalos del 15% de verosimilitud para el parámetro umbral μ utilizando $R_r^p(\mu)$ y $R_D^p(\mu)$ son (3.19, 8.25) y (3.19, 8.31), respectivamente. En ambos casos el emv es 7.80. El intervalo del 5% de verosimilitud obtenido con $R_D^p(\mu)$ es (-1.68, 8.33), el cual incluye algunos valores negativos para μ . Sin embargo, para el valor de $\mu = -1.68$ se tienen las estimaciones restringidas de $\hat{\sigma}(\mu) = 16.21$ y $\hat{\beta}(\mu) = 5.08$ y el modelo Weibull que correspondería a estos parámetros asigna una probabilidad de 0.00001 al evento de observar diámetros negativos, la cual es despreciable. En las simulaciones efectuadas con un modelo Weibull con parámetro de forma $1 < \beta \leq 2$, las coberturas de los parámetros μ , σ y β fueron cercanas a 0.9,

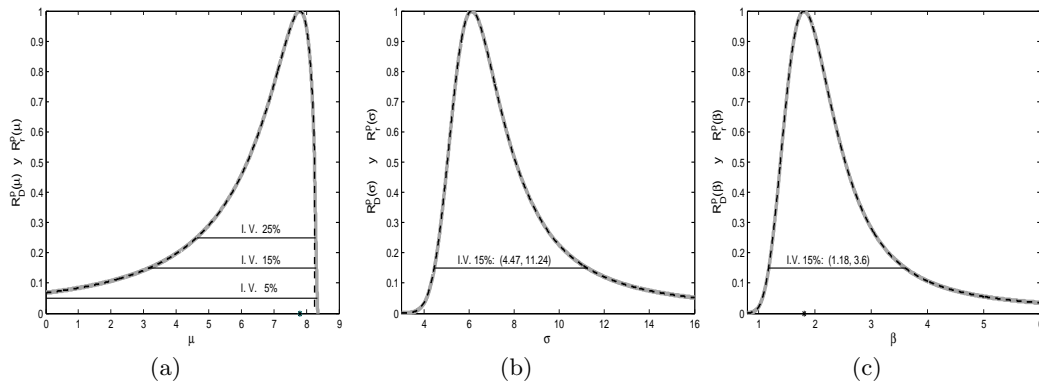


Figura 4.1: Diámetros de árboles: Verosimilitudes relativa perfil restringida R_r^p (---) y discretizada R_D^p (—) de los parámetros (a) μ , (b) σ y (c) β .

cuando la precisión utilizada fue de $2h = 0.01$; por tanto coberturas similares serían las que se estarían asociadas a los intervalos del 15% de verosimilitud del umbral. La Figura 4.2(a) muestra la gráfica cuantil-cuantil con el modelo Weibull estimado, y con base en ella se puede considerar razonable este modelo para describir a los diámetros de árboles de cerezas negras.

Ahora, analizando la verosimilitud perfil relativa para el parámetro de forma c de la DGVE, que se muestra en la Figura 4.2(b), puede observarse que la muestra favorece solamente valores negativos de c , lo que nuevamente sustenta un modelo Weibull. El valor de $c = 0$, asociado al modelo Gumbel, tiene una verosimilitud relativa despreciable.

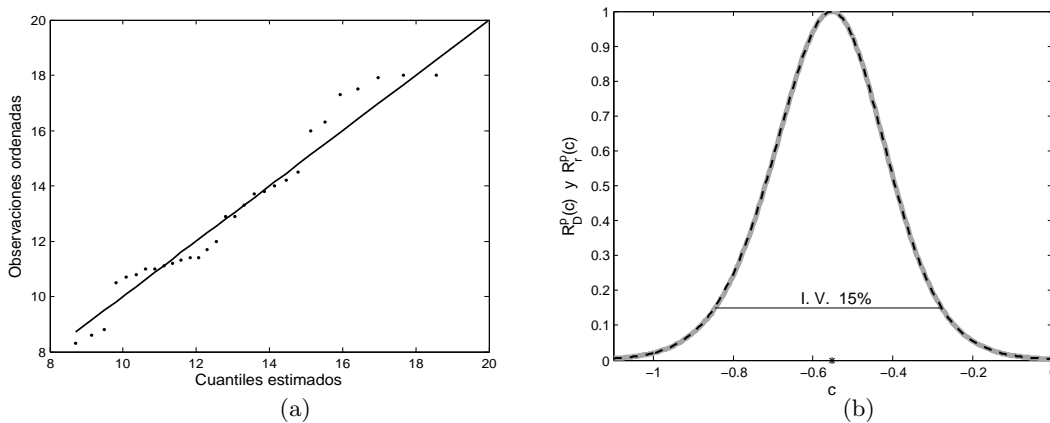


Figura 4.2: Datos de árboles: (a) Gráfica cuantil-cuantil bajo el modelo Weibull estimado. (b) Verosimilitud relativa perfil del parámetro de forma c de la DGVE, $R_r^p(c)$ (---), $R_D^p(c)$ (—).

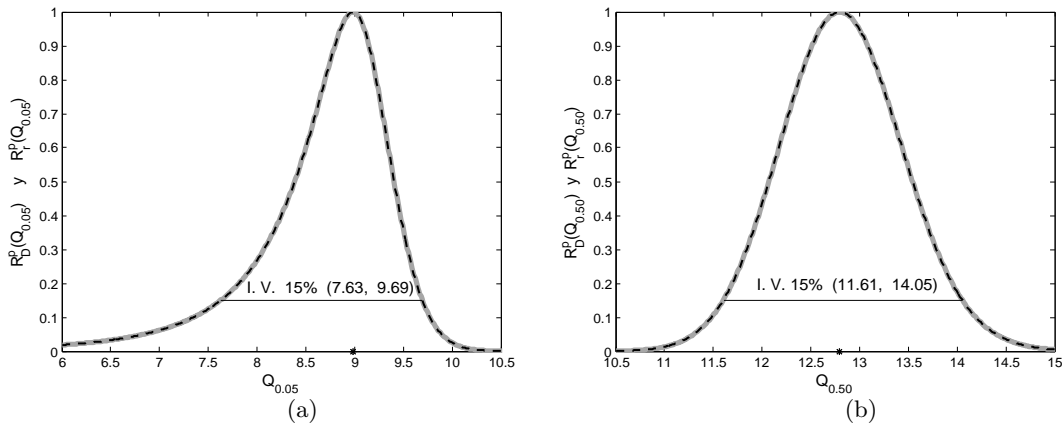


Figura 4.3: Diámetros de árboles: Verosimilitudes relativa perfil restringida R_r^p , (---) y discretizada R_D^p (—) de los cuantiles (a) $Q_{0.05}$ y (b) $Q_{0.50}$.

Un cuantil pequeño como $Q_{0.05}$, donde aproximadamente el 5% de las observaciones caen a la izquierda de este cuantil, también da información parecida a la del umbral, que representa una cota inferior de los datos. La verosimilitud relativa perfil de $Q_{0.05}$, se muestra en la Figura 4.3(a). El intervalo del 15% de verosimilitud es (7.63, 9.69). Esto se puede interpretar como que aproximadamente el 95% de estos cerezos tiene un diámetro de por lo menos 7.63 pulgadas, o que hay aproximadamente un 5% de cerezos cuyo diámetro es cuando mucho de 9.69 pulgadas.

En el Escenario 2 de simulaciones de la Sección 3.6 se utilizaron valores de parámetros similares a las estimaciones obtenidas con estos datos. Las coberturas de los intervalos de verosimilitud del parámetro umbral y del cuantil $Q_{0.05}$ fueron buenas y muy similares entre sí, con valores cercanos a 0.91, desde muestras de tamaño 25. Con base en esto, se pueden considerar razonables los intervalos de verosimilitud calculados en este ejemplo.

Ahora, la verosimilitud perfil de la mediana $Q_{0.50}$ que se muestra en la Figura 4.3(b), también es muy informativa, pues proporciona información sobre un valor central de la distribución de estos diámetros. El intervalo del 15% de verosimilitud de $Q_{0.50}$ es (11.61, 14.05) y se muestra en esta figura. La curva es mucho más simétrica que la verosimilitud perfil de un cuantil pequeño. Esto suele ocurrir en general (Sprott, 2000, p.184), como se podrá ver en otros de los ejemplos de este capítulo.

4.1.2 Tiempos de detección de cáncer en ratas experimentales

El cáncer es una de las enfermedades más estudiadas y para la cual se asocian diversos factores de riesgo. Muchos experimentos, como el realizado por Pike (1966) y que presenta Lawless (2003, p. 188), son estudios relacionados sobre inducción de

cáncer. En dicho experimento se pintó la vagina de 19 ratas de laboratorio con un cancerígeno y se procedió a contar el número T de días, que pasaron hasta poder detectar un carcinoma. Estos datos se presentan en la Sección C.2, del Apéndice C y se puede considerar que el error del instrumento de medición es $2h = 1$ día, ya que se registra el día más cercano en que se observa el carcinoma. Dos de las ratas dieron lugar a observaciones censuradas, pues solamente se sabe que hasta el día registrado no se les había detectado carcinoma alguno. Considerar datos censurados es muy sencillo bajo el enfoque de verosimilitud, como se explicó en la Sección 1.3.6 de esta tesis.

Lawless analiza estos datos con un modelo Weibull de tres parámetros y utilizando la aproximación continua a la verosimilitud. Aquí se utilizan las verosimilitudes restringida y discretizada para estimar los parámetros de este modelo y se obtiene $\hat{\mu} = 122.05$, $\hat{\sigma} = 108.41$ y $\hat{\beta} = 2.71$, con ambas verosimilitudes. Estos resultados son muy similares a los obtenidos por Lawless.

La Figura 4.4(a) muestra la verosimilitud perfil del parámetro umbral μ , en la que puede observarse existe un máximo local, con una estimación máximo verosímil de μ distinta de cero. El intervalo del 15% de verosimilitud para μ es $(-541.5, 142.4)$, donde el valor de $\mu = -541.5$ genera las estimaciones restringidas $\hat{\sigma}(\mu) = 778.14$ y $\hat{\beta}(\mu) = 20.01$ y la probabilidad de observar valores negativos bajo un modelo Weibull con estos valores de parámetros es despreciable, pues resulta 0.0007. De acuerdo a las simulaciones realizadas, cuando el parámetro de forma es $\beta \geq 2$, las coberturas de los intervalos de verosimilitud perfil de modelo Weibull, son mayores que 0.9, aún en muestras pequeñas, por lo que es razonable proporcionar estos intervalos de estimación.

Cuando se analizan estos datos utilizando la DGVE, el emv del parámetro de forma c de esta distribución, resulta $\hat{c} = -0.37$ y su verosimilitud perfil se muestra en la Figura 4.4(b), donde es claro que los datos favorecen valores negativos de c y por lo tanto la selección de un modelo Weibull parece adecuada, como también se corrobora más adelante con las gráficas cuantil-cuantil que se presentan en la Figura 4.8.

Debido a que el valor $\mu = 0$ tiene una verosimilitud relativa moderada, de aproximadamente 0.4, Lawless sugiere la posibilidad de utilizar un modelo Weibull de dos parámetros, que resulta regular y para el cual se obtuvieron las estimaciones $\hat{\sigma} = 234.37$ y $\hat{\beta} = 6.08$.

Se comparará a estos dos modelos a través de las verosimilitudes perfiles de ciertos cuantiles. Lawless no realiza esta comparación y prefiere el modelo Weibull de dos parámetros por ser razonable para los datos y regular, con lo que se simplifican mucho las estimaciones. También menciona en su libro que la estimación de los cuantiles es muy estable con ambos modelos.

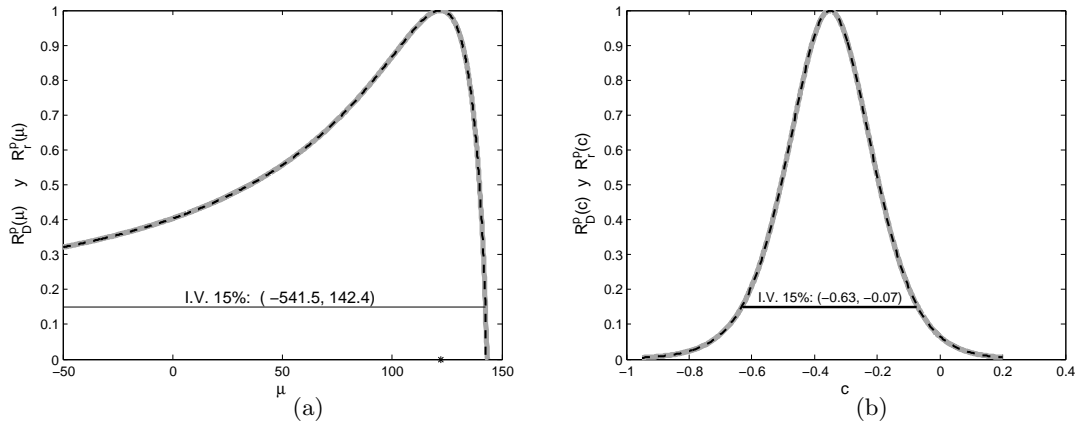


Figura 4.4: Datos de Pike: (a) Verosimilitud relativa perfil restringida $R_r^p(\mu)$, (---) y discretizada $R_D^p(\mu)$ (—) (b) Verosimilitud relativa perfil del parámetro de forma c de la DGVE, $R_r^p(c)$ (---), $R_D^p(c)$ (—).

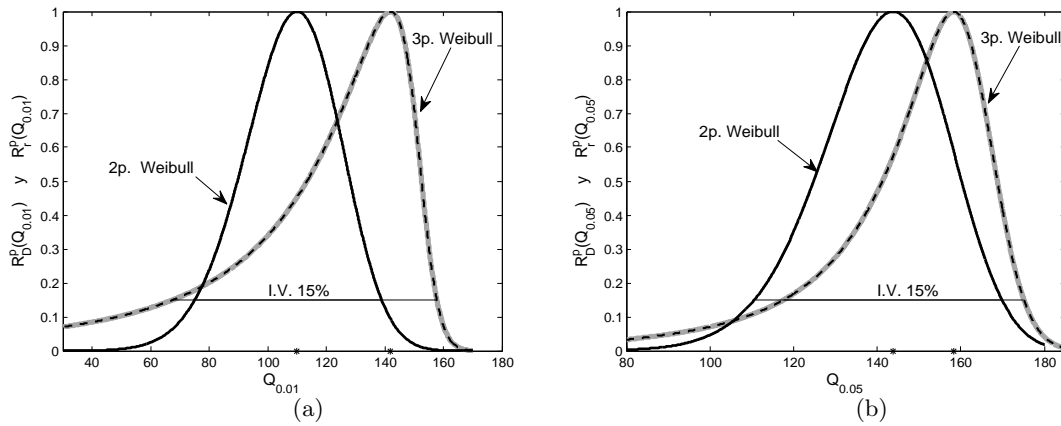


Figura 4.5: Datos de Pike: Verosimilitud relativa perfil restringida, R_r^p (---) y discretizada R_D^p (—) para los cuantiles (a) $Q_{0.01}$ y (b) $Q_{0.05}$.

Aquí se considerará la estimación de cuantiles pequeños, como $Q_{0.01}$ y $Q_{0.05}$, que pueden ser una estimación alternativa al umbral, pero también se calculan las verosimilitudes del cuantil $Q_{0.50}$, que proporciona información de un valor central de la distribución y de un cuantil grande como Q_{99} , que también es informativo. Analizando las verosimilitudes perfiles de estos cuantiles se podrá evaluar la robusticidad de las inferencias de estos parámetros de localización al usar los dos modelos considerados.

Las verosimilitudes perfiles de $Q_{0.01}$ y $Q_{0.05}$ se muestran en la Figura 4.5. Nótese que las curvas alcanzan el máximo en lugares distintos, pero se traslapan en gran medida. Los intervalos del 15% de verosimilitud coinciden en gran parte. En el caso de $Q_{0.01}$ son más anchos los intervalos del modelo Weibull de tres parámetros e incluyen a los del modelo de dos parámetros. Esto ocurre porque el modelo de tres

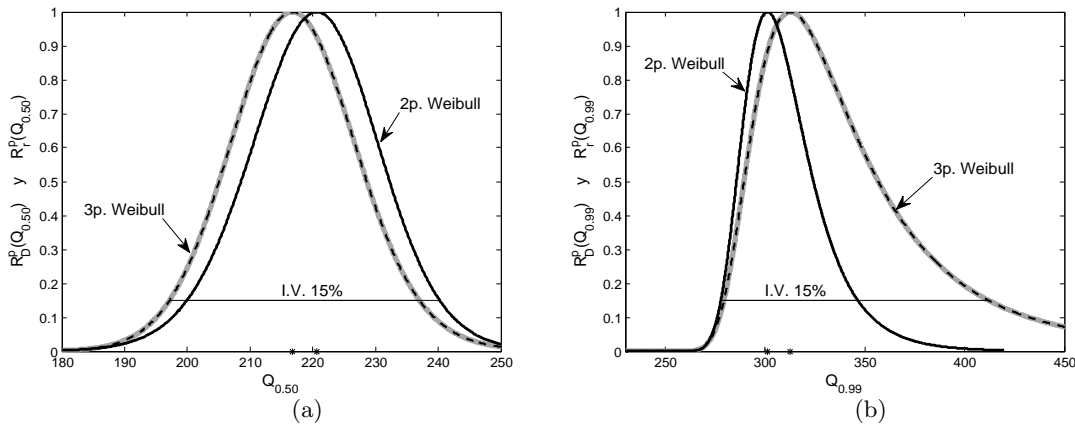


Figura 4.6: Datos de Pike: Verosimilitud relativa perfil restringida, R_r^p (---) y discretizada R_D^p (—) para los cuantiles (a) $Q_{0.50}$ y (b) $Q_{0.99}$.

parámetros contempla la incertidumbre que se tiene sobre el umbral y resulta una alternativa conservadora para adoptar.

En el caso de $Q_{0.50}$, la Figura 4.6(a) muestra que las inferencias que se obtienen con ambos modelos son casi las mismas, y además la forma de la verosimilitud se simetriza notoriamente. Para estimar la mediana se puede adoptar el modelo más simple de dos parámetros sin mayor problema o consideración, sujeto a que se vea un buen ajuste a los datos a través de la gráfica cuantil-cuantil, como sí se observa en la Figura 4.8(b). Sin embargo, nuevamente vuelve a haber discrepancias en las verosimilitudes perfiles del cuantil grande $Q_{0.99}$, debido a la incertidumbre que sobre el umbral se tiene con estos datos. En este caso sería mejor conservar el modelo Weibull de tres parámetros cuyos intervalos de estimación incluyen a los que se obtienen con el modelo de dos parámetros.

Ahora, con respecto a la estimación de $Q_{0.05}$, las simulaciones mostraron que en estos casos sus intervalos de verosimilitud perfil construidos bajo un modelo Weibull de tres parámetros pueden tener una cobertura mayor de 0.9, en muestras de tamaño 25, por lo que los intervalos de estimación con este modelo para los cuantiles sí son razonables para el ejemplo en consideración.

En cuanto a la robusticidad de las inferencias, se tiene un panorama muy distinto al considerar estos dos modelos y estimar los parámetros de forma y de escala. Las verosimilitudes perfiles se muestran en la Figura 4.7, y se nota claramente las discrepancias. Es importante notar que a pesar de las diferencias que se observaron en las verosimilitudes perfiles de los cuantiles, éstas no son tan marcadas como las que se observan en la estimación de σ y β . Esto es, en cuanto a la estimación de los cuantiles, los dos modelos dan inferencias mucho más robustas y estables.

En cuanto al ajuste de los dos modelos a los datos, las gráficas cuantil-cuantil

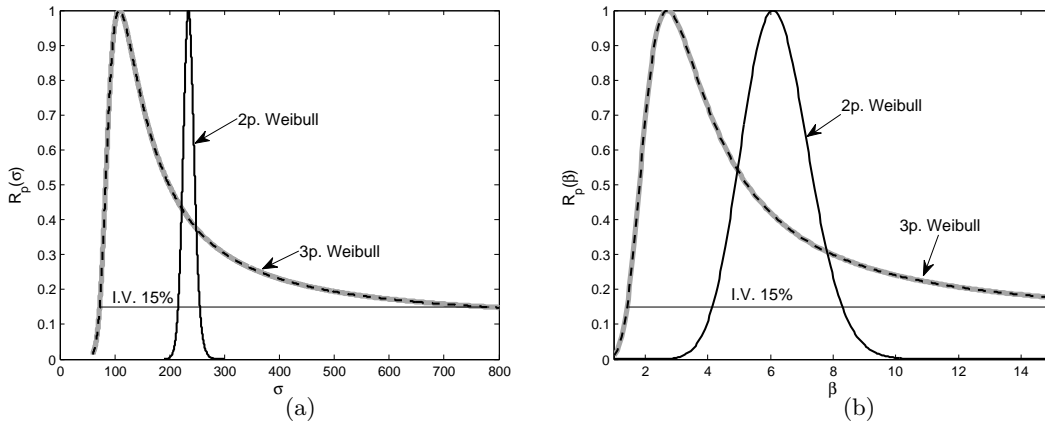


Figura 4.7: Datos de Pike: Verosimilitud relativa perfil de los parámetros (a) σ y (b) β , bajo los modelos Weibull de tres y dos parámetros.

dados en la Figura 4.8(a) y 4.8(b) muestran que ambos modelos serían razonables para los datos.

En resumen, si el interés reside en estimar la mediana o algún cuantil central, se

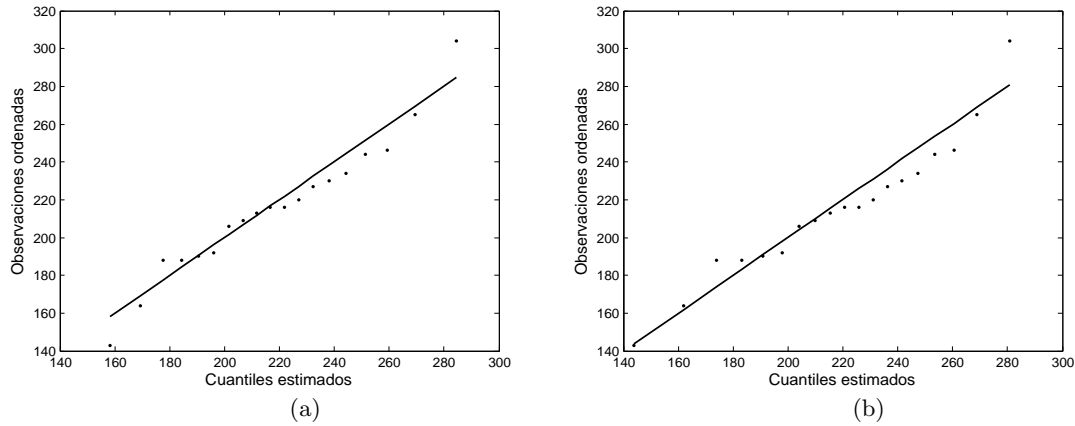


Figura 4.8: Datos de Pike: (a) Gráfica cuantil-cuantil para los modelos Weibull estimados: (a) De tres parámetros (b) De dos parámetros.

recomendaría seleccionar el modelo Weibull de dos parámetros por ser más simple y regular. Sin embargo, si se desea estimar cuantiles en las colas de la distribución o el parámetro umbral, se debiera utilizar un modelo Weibull de tres parámetros, cuyas inferencias son mucho más conservadoras, además de tener buenas coberturas. Aún para muestras con $n = 25$, los intervalos de verosimilitud del umbral mostraron tener coberturas razonables, mayores al 90%, en el Escenario 2 simulado con este modelo. Una comparación de las estimaciones puntuales y por intervalo que se obtienen para varios cuantiles y los parámetros σ y β de la distribución Weibull de dos y tres parámetros, se muestra en la Tabla 4.1.

Parámetro	<i>Weibull: dos parámetros</i>		<i>Weibull: tres parámetros</i>	
	emv	15% I. V.	emv	15% I. V.
$Q_{0.01}$	108.41	(75.1,138.9)	141.89	(67.4,157.7)
$Q_{0.05}$	143.80	(110.6,169.8)	158.26	(117.5,175.1)
$Q_{0.50}$	220.66	(200,240.2)	216.74	(197.2,237)
$Q_{0.99}$	301.19	(277.9,346.4)	312.55	(279.4,411.8)
σ	234.37	(215.3,255.1)	108.41	(72.4,781.9)
β	6.08	(4.16,8.26)	2.71	(1.46, 19.96)

Tabla 4.1: Datos de Pike: Estimaciones puntuales e intervalos del 15% de verosimilitud bajo los modelos estimados Weibull de dos y tres parámetros.

En todas las gráficas se presentaron tanto la verosimilitud perfil discretizada como la restringida. Puede observarse que sus curvas se traslapan completamente y son siempre iguales, para este ejemplo. En particular, este ejemplo muestra cómo las gráficas de las verosimilitudes perfiles de parámetros de interés son muy útiles para evaluar la robusticidad de sus inferencias, bajo distintos modelos.

4.1.3 Índice de masa corporal en mujeres sonorenses

Se analizaron los datos de altura y de peso de 552 mujeres sonorenses, no embarazadas, seleccionados de una muestra aleatoria de mujeres mexicanas, representativa por estado, la cual se obtuvo de la Encuesta Nacional de Salud y Nutrición (ENSANUT), que efectuó el Instituto Nacional de Salud Pública, en el año 2006. La precisión en las mediciones de altura T , medida en metros, es de $2h_t = 0.001$ y para el peso W , que fue medido en kilogramos, se tiene una precisión de $2h_w = 0.1$.

En este problema, la variable de interés es el índice de masa corporal (IMC) el cual se utiliza tanto para medir desnutrición, sobrepeso y obesidad, como para evaluar riesgos cardiovasculares, al ser proporcional al porcentaje de grasa corporal de un individuo. El índice de masa corporal IMC, que denotaremos por Z , se define como el cociente del peso de una persona, en kilogramos, dividido entre el cuadrado de la altura, en metros

$$Z = \frac{W}{T^2}. \quad (4.1.1)$$

Actualmente, este índice se utiliza a nivel mundial como una medida sencilla, rápida y barata, de predecir el porcentaje de grasa corporal en el cuerpo y con base en éste se pueden evaluar posibles riesgos cardiovasculares. Así, si $Z \leq 18$ se dice que la persona está desnutrida; si $18 < Z \leq 25$, la persona tiene un peso normal; si $25 < Z \leq 28$ la persona tiene sobrepeso y finalmente, si $Z > 28$ la persona es obesa. Se suelen hacer ajustes a estos criterios según sea el género, la edad y la actividad física. Los límites que se mencionan, se fijaron considerando los cuantiles de una muestra grande tomada de la población caucásica en Estados Unidos, en la década de 1980. El valor de $Z = 25$ correspondía a la mediana de los IMC de esa muestra

y el valor $Z = 28$ representaba el cuantil $Q_{0.85}$.

En el ejemplo a analizar, la observación mínima para el IMC fue de $z_{(1)} = 15.182$, la máxima $z_{(552)} = 56.112$, mientras que la media y varianza muestrales fueron $\bar{z} = 29.87$ y $s^2 = 36.57$, respectivamente. Una característica interesante en este problema es que se conoce la precisión de la altura T , medida en metros, $2h_t = 0.001$ y del peso W medido en kilogramos, $2h_w = 0.1$, pero no se tiene una medición directa del IMC. En este ejemplo la precisión del IMC está en función de otras precisiones, por lo que su error de medición es indirecto. Esta situación es bien conocida en física y se le llama propagación de errores (Shibata et al., 1995, p. 34-36) y de acuerdo con ella la precisión correspondiente a Z es una función de h_t , h_w , T , y W . Es por ello que cada una de las mediciones Z_i puede tener una precisión diferente $2h_{z_i}$. Esta es una característica que en ocasiones tienen los datos y que generalmente no se toma en cuenta, pues muchas técnicas de estimación no permiten considerarla. Sin embargo, la verosimilitud discretizada, presentada en la Definición 3.2, permite incorporar esta característica de una manera muy sencilla y natural.

La densidad normal es simétrica y por tanto no es razonable para describir este conjunto de datos, ya que la distribución de éstos tiene una cola derecha muy pesada, como se puede observar en el histograma de estos índices, mostrado en la Figura 4.9(a). Por ello, un modelo lognormal de tres parámetros, con la densidad dada en (2.5.3) parece ser más adecuado para modelar el índice de masa corporal. Las estimaciones puntuales que se obtienen bajo este modelo son $\hat{\mu} = 1.7686$, $\hat{\gamma} = 3.3131$ y $\hat{\sigma} = 0.2124$, densidad que se dibuja encima del histograma presentado en la Figura 4.9(a). La Figura 4.9(b) muestra la gráfica cuantil-cuantil de los datos, usando estas estimaciones y el ajuste resulta bastante bueno.

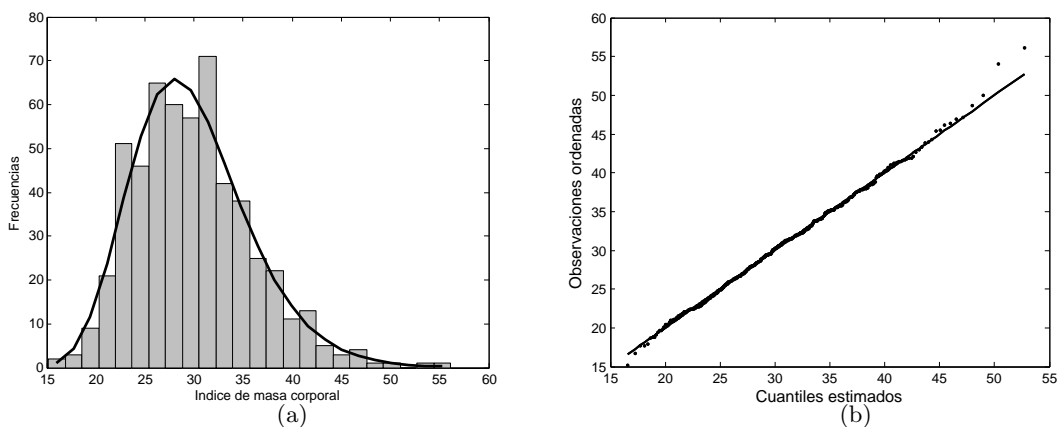


Figura 4.9: Datos de IMC: (a) Histograma del IMC, (b) Gráfica cuantil-cuantil con el modelo lognormal de tres parámetros estimado.

Bajo este modelo, la verosimilitud perfil del parámetro umbral μ , obtenida tanto con la verosimilitud restringida como con la verosimilitud discretizada, proporci-

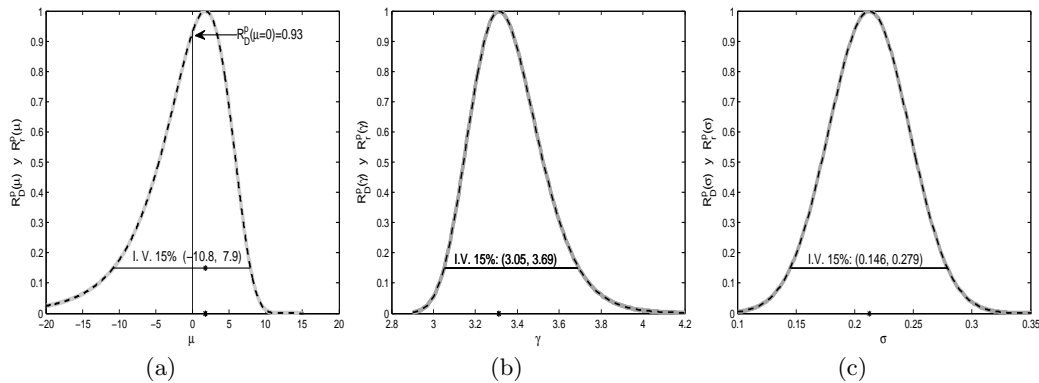


Figura 4.10: Datos de IMC: Verosimilitudes perfil relativa restringida R_r^p , (---) y discretizada R_D^p (—) de (b) μ , (b) γ , (c) σ , bajo un modelo lognormal.

nan inferencias similares, como puede verse en la Figura 4.10(a). Lo mismo puede observarse en las verosimilitudes perfil de γ y de σ , los otros parámetros del modelo, que se muestran en las Figuras 4.10(b) y 4.10(c), en las que se indican sus intervalos del 15% de verosimilitud.

En la Figura 4.10(a) puede observarse que utilizar un modelo lognormal de dos parámetros podría ser una alternativa más simple para modelar estos datos, ya que $R_D^p(\mu = 0) = 0.93$, esto es, la plausibilidad de $\mu = 0$ es alta. Una forma de verificar esto es graficar, para ambos modelos, las verosimilitudes perfiles de ciertos cuantiles de interés. Esto se muestra en la Figura 4.11.

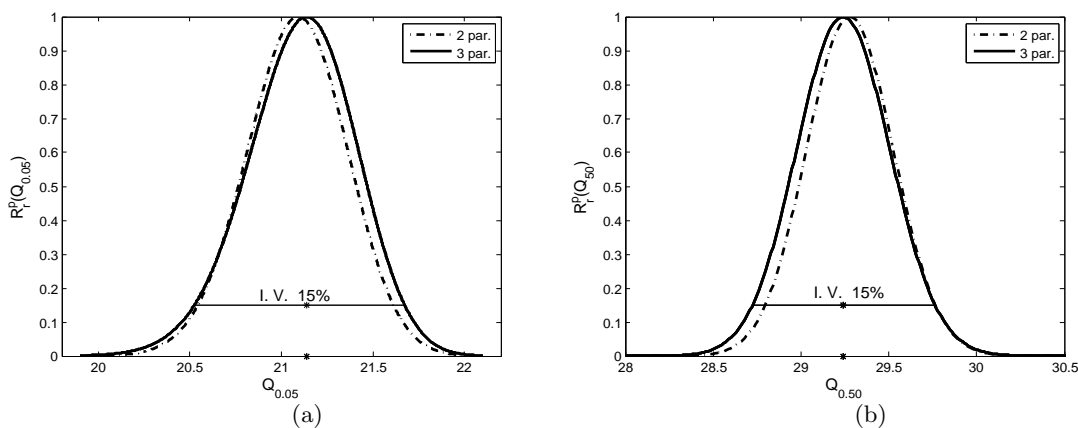


Figura 4.11: Datos de IMC: Verosimilitudes perfil relativa de algunos cuantiles (a) $Q_{0.05}$ y (b) $Q_{0.50}$.

En este problema, la verosimilitud perfil de un cuantil pequeño, como lo es $Q_{0.05}$, permite detectar posibles problemas de desnutrición. Como puede apreciarse en la

Figura 4.11(a) donde se muestra la verosimilitud perfil para este cuantil, las mujeres sonorenses no tienen problemas de desnutrición, ya que todos los valores para este cuantil son mayores a 20. Por otra parte, la mediana también es un parámetro muy informativo. En la Figura 4.11(b) se muestra la verosimilitud perfil de $Q_{0.50}$, donde se observa que toma valores mayores a 28.5. Por tanto, hay evidencia fuerte de que más de la mitad de las mujeres sonorenses presentan sobrepeso u obesidad.

Para estos datos, las inferencias sobre cuantiles hechas con los dos modelos son muy robustas. Por ello, será preferible el modelo más sencillo, la lognormal de dos parámetros, ya que además ajusta muy bien a los datos, como se muestra en la gráfica cuantil-cuantil de la Figura 4.12(a).

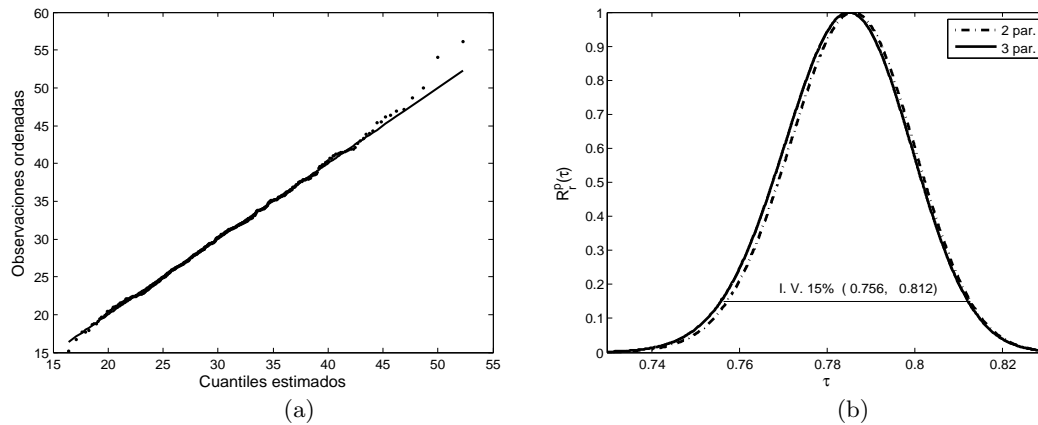


Figura 4.12: Datos de IMC: (a) Gráfica cuantil-cuantil con un modelo lognormal de dos parámetros (b) Verosimilitud perfil relativa del parámetro $\tau = P(Z > 25)$.

En este estudio resulta de gran interés conocer la probabilidad de que una mujer sonorense tenga sobrepeso u obesidad. El parámetro que proporciona esta información sería $\tau = P[Z > 25]$. La verosimilitud perfil relativa de este parámetro puede obtenerse fácilmente gracias a la propiedad de invarianza funcional de la verosimilitud, frente a reparametrizaciones. La gráfica de la verosimilitud perfil relativa restringida de τ , se muestra en la Figura 4.12(b), en la cual se utiliza tanto un modelo lognormal de tres, como de dos parámetros. Puede observarse que los intervalos de verosimilitud perfil del 15% son prácticamente los mismos e indican que entre el 76% y 81% de las mujeres sonorenses tienen problemas de obesidad, lo cual indica un serio problema de salud.

Con lo anterior, se vuelve a mostrar cómo la forma y en sí toda la verosimilitud perfil de un parámetro de interés, es de utilidad tanto para seleccionar un modelo, como para dar respuesta a una pregunta específica de interés.

4.2 Ejemplos tipo II

4.2.1 Tiempos de espera en ratas para realizar una tarea de aprendizaje

Se condujo un experimento para explorar y evaluar el efecto que una droga, supuestamente amnésica, tiene en la memoria de ratas experimentales y en sus procesos de aprendizaje. Para ello, el grupo de investigación dirigido por el Dr. Roberto Prado-Alcalá, del Instituto de Neurobiología de la UNAM, campus Querétaro, efectuó el experimento que se describe enseguida.

Se asignaron 24 ratas blancas, al azar, a dos grupos. El primer grupo de doce ratas fue inyectado con la droga amnésica y el otro grupo, también de doce ratas, fue inyectado con un placebo. En el experimento se cuenta con una caja rectangular,

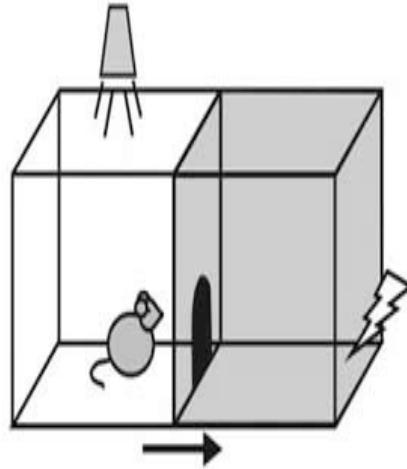


Figura 4.13: Caja experimental para medir los tiempos de latencia de retención.

dividida en dos compartimentos de forma cuadrada. Uno de estos compartimentos está iluminado y el otro está oscuro, como se muestra en la Figura 4.13. El experimento consiste en poner a cada rata, una por vez, en el centro de la parte iluminada de la caja. En uno de los lados de este compartimento iluminado hay una puerta que conduce al compartimento oscuro. Como las ratas prefieren la oscuridad, tienden a pasar a ese compartimento, en cuanto encuentran la puerta por la que pueden hacerlo. Cuando entran al lado oscuro, se cierra una compuerta que separa los dos compartimentos y durante cinco segundos la rata recibe un choque eléctrico. Después de este tiempo, se vuelve a abrir la compuerta y se mide el tiempo que tarda la rata en escapar. Al día siguiente, se coloca a la misma rata nuevamente en el compartimento iluminado y se mide el tiempo que tarda en entrar al compartimento oscuro. Estos tiempos, llamados latencias de retención, son altamente informativos en cuanto a la relación entre la droga de interés (por su posible efecto amnésico) y el proceso de aprendizaje. Las ratas que recuerden bien el choque eléctrico del día anterior, evitarán pasar al compartimento obscuro ó tardarán más tiempo antes de entrar en éste, a diferencia de aquellas ratas que hayan olvidado el choque, debido a la droga que se les inyectó.

En la Sección C.3 del Apéndice C, se proporcionan algunas medidas descriptivas correspondientes a las mediciones de estos tiempos, tanto para el grupo de 12 ratas a las que se les administró placebo, como para las otras 12 que estuvieron bajo el efecto de la droga, ambos grupos recibiendo un choque eléctrico de 0.8 mA. Los tiempos fueron medidos con un cronómetro que tiene una precisión de $2h = 0.01$ segundos.

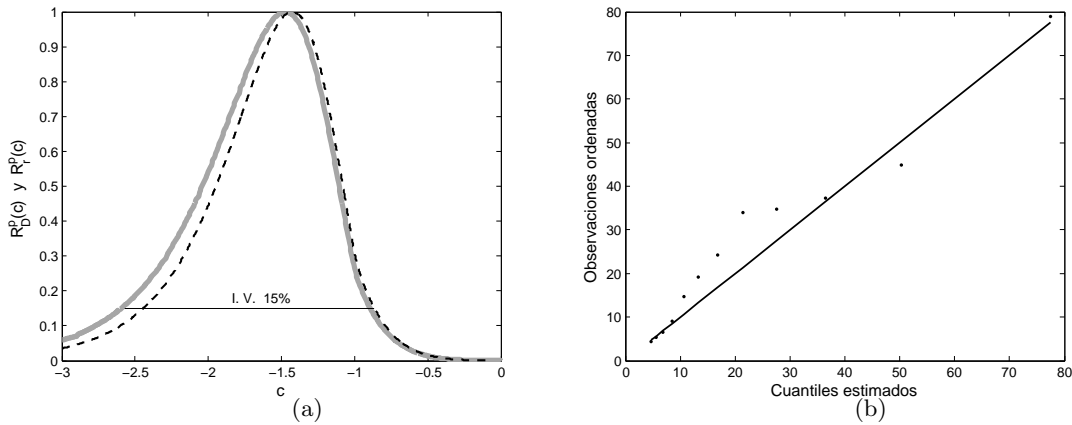


Figura 4.14: Tiempos de espera de las ratas en el grupo placebo: (a) Verosimilitud relativa perfil del parámetro c de la DGVE; verosimilitud restringida $R_r^p(c)$ (---) y verosimilitud discretizada $R_D^p(c)$ (—). (b) Gráfica cuantil-cuantil para el modelo Weibull de tres parámetros estimado.

Dadas las características del problema, el modelo Weibull de tres parámetros podría ser considerado para analizar estos datos. Esta selección puede reafirmarse construyendo la verosimilitud perfil del parámetro de forma c de la distribución de valores extremos generalizada, que como puede verse en la Figura 4.14(a), solamente toma valores negativos, los cuales sustentan a un modelo Weibull. En esta figura se utilizó tanto la verosimilitud perfil restringida, como la discretizada y ambas sugieren este modelo como adecuado para describir los datos. Aunque no se muestra gráficamente, lo mismo ocurre cuando se analiza el conjunto de datos de los tiempos medidos en las ratas que estuvieron bajo el efecto de la droga.

Los tiempos de retención, para las ratas que tomaron placebo, dió lugar a una verosimilitud perfil del umbral en forma de J . Este es un ejemplo donde no es posible la estimación por máxima verosimilitud cuando se utiliza la aproximación continua L_C , ya que la verosimilitud no es acotada, no tiene un máximo local y posee una singularidad en $\mu = x_{(1)}$. La verosimilitud perfil relativa restringida de μ , $R_r^p(\mu)$ tiene también forma de J , pero es acotada; el máximo se alcanza en $\hat{\mu} = x_{(1)} - h = 4.32 - 0.005 = 4.315$ y los emv de σ y β son 18.44 y 0.7012, respectivamente. La verosimilitud perfil relativa discretizada para este parámetro umbral, $R_D^p(\mu)$, es también acotada y unimodal, con un pico pronunciado que alcanza su

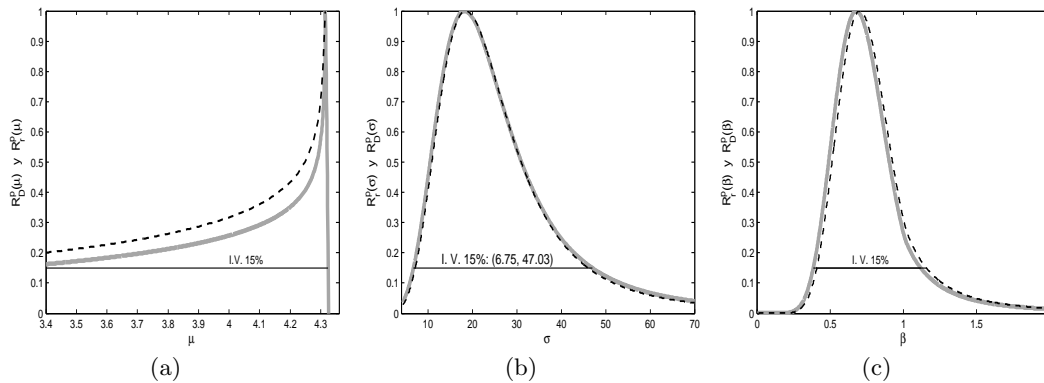


Figura 4.15: Tiempos de espera en experimento de ratas en el grupo placebo. Verosimilitud relativa perfil restringida R_r^p (---) y discretizada R_D^p (—) de los parámetros: (a) μ , (b) σ y (c) β del modelo Weibull.

máximo en este mismo valor $\hat{\mu}$ y después decrece a cero en $x_{(1)} + h/2$. Los emv que se alcanzaron con L_D , para σ y β fueron $\hat{\sigma} = 18.3$, y $\hat{\beta} = 0.68$, para el grupo donde se administró placebo. El hecho que $\hat{\beta} < 1$, favorece una densidad Weibull convexa, y por tanto este modelo estimado asigna probabilidades altas a las observaciones cercanas al umbral.

En la Figura 4.15(a) se muestran las verosimilitudes perfil de μ obtenidas con la verosimilitud restringida y con la verosimilitud discretizada. Puede verse que la verosimilitud restringida se sitúa por arriba de la discretizada, por lo que sus intervalos de verosimilitud serán más anchos e incluirán por completo a los de la discretizada. El intervalo del 15% de verosimilitud para μ , con la verosimilitud discretizada es (3.257, 4.324), mientras que con la verosimilitud restringida resulta (2.80, 4.31). Este es el único caso donde se ha visto discrepancias importantes entre las dos verosimilitudes. La plausibilidad de $\mu = 0$, que favorecería un modelo Weibull de dos parámetros es pequeña con ambas verosimilitudes, $R_D^p(\mu = 0) = 0.05$ y $R_r^p(\mu = 0) = 0.06$, por lo que el modelo Weibull de tres parámetros seleccionado, es razonable, como se muestra en la gráfica cuantil-cuantil dada en la Figura 4.14(b), para el grupo placebo.

La verosimilitud perfil de σ se muestra en la Figura 4.15(b) y su intervalo del 15% de verosimilitud es (6.75, 47.03), tanto con la verosimilitud restringida como con la verosimilitud discretizada. El intervalo de verosimilitud del 15% para β se muestra en la Figura 4.15(c), resultando (0.39, 1.12) con la verosimilitud discretizada y (0.41, 1.15), con la verosimilitud restringida.

De acuerdo a las simulaciones efectuadas en el Capítulo 3, cuando se estima un parámetro umbral en condiciones como las aquí presentadas, lo cual fue simulado en el Escenario 1, las coberturas de los intervalos de verosimilitud perfil del um-

bral no son buenas, ya que andan por debajo de 0.7, tanto para la verosimilitud restringida como para la discretizada. Por otra parte, en la estimación de un cuantil pequeño, como $Q_{0.05}$, las coberturas oscilan entre 0.7 y 0.8, cuando las muestras son de tamaño 25, por lo que en este caso pudieran ser aún menores. Este es un caso donde sería mejor estimar un cuantil pequeño en vez del umbral, ya que la muestra es demasiado pequeña para poder dar inferencias razonables para el umbral. Por otra parte, las coberturas de los intervalos de verosimilitud de los parámetros σ y β , estuvieron cercanas a los valores teóricos, por lo que resulta razonable proporcionar dichos intervalos para esos parámetros, con el tamaño de muestra con que se cuenta en este ejemplo.

Con respecto al grupo de ratas al que se le administró la droga que induce pérdida de memoria, el modelo que resultó razonable utilizar fue también un modelo Weibull de tres parámetros, para el cual se encuentra un máximo local con la aproximación continua a la verosimilitud L_C . Las estimaciones de máxima verosimilitud que se obtienen con la verosimilitud restringida L_r y la verosimilitud discretizada L_D , son prácticamente las mismas que se obtienen con L_C y son $\hat{\mu} = -1.76$, $\hat{\sigma} = 36.9$ y $\hat{\beta} = 1.78$.

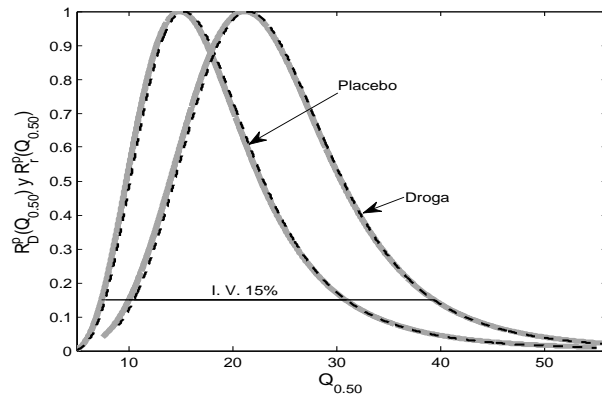


Figura 4.16: Tiempos de entrenamiento en las ratas en los grupos bajo la droga y el placebo; verosimilitudes relativas perfil de la mediana $Q_{0.50}$. Verosimilitud restringida (---), y verosimilitud discretizada (—).

La comparación entre las medianas de estos grupos es de primordial importancia para los investigadores. Las verosimilitudes relativas perfil restringida y discretizada de la mediana, en ambos grupos de ratas, se presentan en la Figura 4.16, donde se observa que ambas curvas se traslapan, en gran medida, al igual que sus intervalos del 15% de verosimilitud. Esto indica que no hay evidencia que favorezca una diferencia entre las medianas. Los resultados concuerdan con la prueba no-paramétrica efectuada por Díaz-Trujillo et al. (2009). Si este traslape no hubiera existido, o fuera mínimo, proporcionaría evidencia fuerte a favor de una diferencia entre las medianas, así como información respecto a la magnitud y signo de esta discrepancia; en contraste, las pruebas no-paramétricas solamente proporcionan evidencia en contra de la igualdad de las medianas.

De acuerdo a las simulaciones efectuadas en el Escenario 1, cuyos parámetros eran similares a los estimados en este ejemplo, las coberturas de los intervalos de verosimilitud perfil discretizados, para un cuantil central como es la mediana $Q_{0.50}$, mostraron coberturas cercanas a las teóricas, aún en muestras de tamaño 25. Por tanto, las inferencias que la verosimilitud discretizada proporciona para este parámetro se pueden considerar razonables.

4.3 Ejemplos tipo III

4.3.1 Máximos anuales del nivel del mar en Port Pirie, Australia

A continuación se analiza una muestra de 65 mediciones de los máximos anuales de las alturas diarias del nivel del mar entre 1923 y 1987, en Port Pirie, Australia, presentado por Coles (2004, p.59). Estas mediciones se presentan en la Sección C.4, del Apéndice C. El dato más pequeño es $x_{(1)} = 3.57$ metros, el más grande $x_{(65)} = 4.69$ y la media y desviación estándar son $\bar{x} = 3.9806$ y $s^2 = 0.2405$, respectivamente. Para estos datos se puede considerar $2h = 0.01$ como la precisión del instrumento de medición.

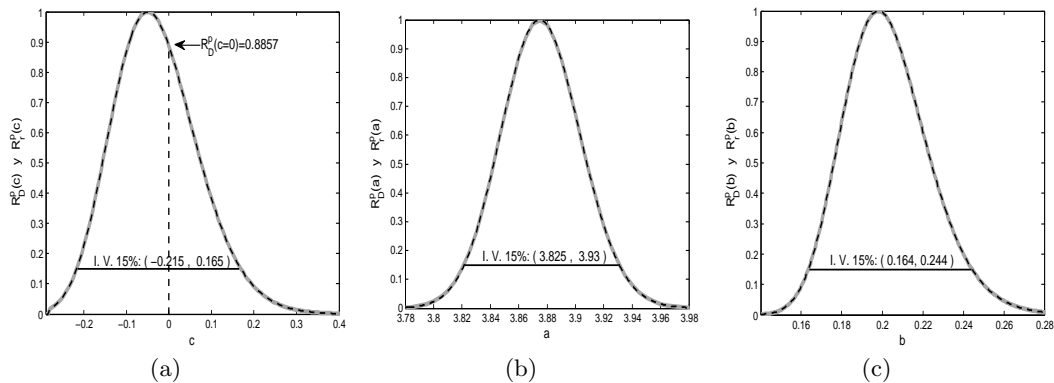


Figura 4.17: Datos de Port Pirie: (a) Verosimilitudes perfil restringida R_r^p (---) y discretizada R_D^p (—) de los parámetros (a) c , (b) a y (c) b de la DGVE.

Coles (2004) consideró razonable suponer que el patrón de variación permaneció constante en el período de observación y propuso modelar los datos como observaciones independientes de la DGVE. Para esta distribución los emv obtenidos fueron $\hat{a} = 3.8748$, $\hat{b} = 0.1981$ y $\hat{c} = -0.0501$. La verosimilitud perfil del parámetro de forma c , que se muestra en la Gráfica 4.17(a), se construyó utilizando tanto la verosimilitud restringida como la verosimilitud discretizada. Con ambas verosimilitudes se obtuvieron las mismas estimaciones puntuales y por intervalo, ya que como puede observarse en esta figura, las curvas se traslapan por completo.

Las verosimilitudes perfil de los parámetros a y b de la DGVE, que se muestran en las Figuras 4.17(b) y 4.17(c) se construyeron también utilizando las verosimilitudes restringida y discretizada y nuevamente son prácticamente indistinguibles. La forma de estas verosimilitudes es similar a la del parámetro c . Este parámetro de forma tiene una verosimilitud relativa alta, igual a 0.89, para el valor $c = 0$ y su intervalo de verosimilitud perfil del 15% toma tanto valores negativos como positivos, lo cual favorece tanto a un modelo Gumbel, como a uno Weibull o a un Fréchet. Como el emv \hat{c} es negativo, se considerará aquí al modelo Weibull, en este análisis, ya que además sirve para ejemplificar la estimación del umbral con las verosimilitudes discretizada y restringidas. Para este conjunto de datos, los parámetros de la DGVE se estimaron fácilmente, gracias a la forma simétrica que tiene la superficie de verosimilitud.

Al ajustar a este conjunto de datos un modelo Weibull de tres parámetros, para máximos, se obtuvieron las estimaciones $\hat{\mu} = 7.83$, $\hat{\sigma} = 3.95$ y $\hat{\beta} = 19.96$. La perfil del umbral μ , bajo este modelo, es plana y extremadamente asimétrica, tanto con la verosimilitud restringida como con la discretizada, las cuales se traslapan por completo. Estas curvas se muestran en la Figura 4.18(a) e indican que valores muy grandes del umbral son razonables para los datos. Recordemos que si el umbral tiende a infinito, es indicativo que se está cerca del modelo límite para la familia Weibull, que sería un modelo Gumbel. Las perfiles para los parámetros σ y β del modelo Weibull de tres parámetros, pueden observarse en las Figuras 4.18(b) y 4.18(c). Éstas se construyeron nuevamente utilizando las verosimilitudes restringida y discretizada y coinciden en todas partes, confirmando la cercanía con el modelo límite Gumbel asociado, ya que son también aplanadas y asimétricas.

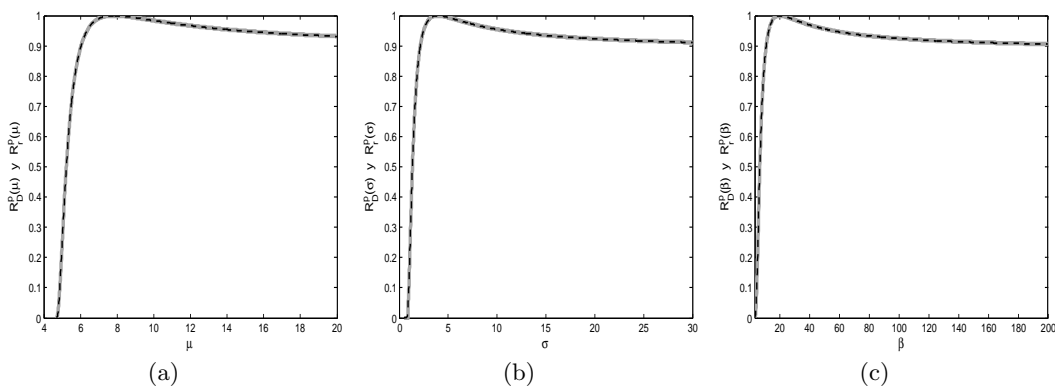


Figura 4.18: Datos de Port Pirie: Verosimilitud perfil restringida R_r^p (---) y discretizada R_D^p (—) de los parámetros (a) μ , (b) σ y (c) β de un modelo Weibull.

Si ahora se ajusta un modelo Gumbel, se obtiene que los emv son 3.87 para el parámetro de localización y 0.19 para el de escala. Dado que interesa estimar una cota superior para estos máximos anuales, un cuantil grande como $Q_{0.95}$ o $Q_{0.99}$, de

este modelo Gumbel, puede servir de alternativa a la estimación de un parámetro umbral. Estos cuantiles también pueden calcularse con los modelos Weibull y DGVE. Se pueden comparar todos estos modelos por medio de las verosimilitudes perfiles de algún cuantil dado. Si se observa robusticidad en las inferencias, entonces cualesquiera de los modelos podría describir adecuadamente estos máximos anuales y en ese caso se seleccionaría al modelo más sencillo.

	<i>I. V. del 15% para $Q_{0.95}$</i>	<i>I. V. del 15% para $Q_{0.99}$</i>
Weibull	(4.31, 4.59)	(4.49, 4.97)
Gumbel	(4.33, 4.60)	(4.60, 4.98)
DGVE	(4.31, 4.66)	(4.49, 5.20)

Tabla 4.2: Datos de Port Pirie: Intervalos del 15% de verosimilitud bajo varios modelos.

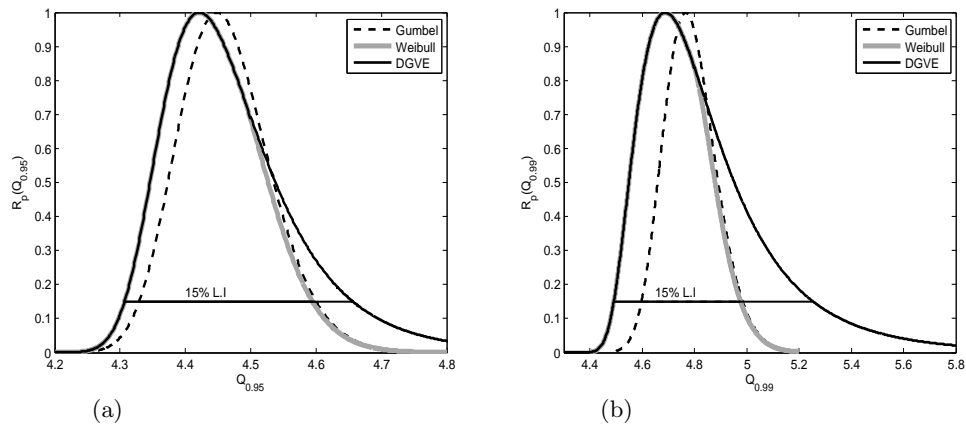


Figura 4.19: Datos de Port Pirie: Verosimilitudes perfil ciertos cuantiles usando los modelos Gumbel, Weibull y la DGVE, (a) $Q_{0.95}$ (b) $Q_{0.99}$

En la Tabla 4.2 se muestran los intervalos del 15% de verosimilitud para los cuantiles $Q_{0.95}$ y $Q_{0.99}$, bajo los modelos Weibull, Gumbel y DGVE. Por otra parte, en la Figura 4.19 se muestra una comparación entre las verosimilitudes perfil para los cuantiles $Q_{0.95}$ y $Q_{0.99}$, bajo estos mismos modelos. En la Figura 4.19(a) puede verse que tanto el modelo Gumbel, como el Weibull, dan inferencias muy similares y son ligeramente diferentes a las de la DGVE. Así, se observa cierta robusticidad en las verosimilitudes perfiles de $Q_{0.95}$, situación que no se tiene en el caso de las verosimilitudes perfil del cuantil $Q_{0.99}$, presentadas en la Figura 4.19(b). En esta puede verse que la verosimilitud perfil de $Q_{0.99}$ bajo el modelo Weibull se traslapa por la izquierda con la obtenida bajo la DGVE y por la derecha con la correspondiente al modelo Gumbel. Algunos valores de $Q_{0.99}$ son plausibles bajo un modelo e implausibles bajo otro, esto es, no se observa robusticidad bajo cambios en el modelo. El intervalo del 15% de verosimilitud que se obtiene con la DGVE, es mucho más ancho que los obtenidos con los modelos Gumbel y Weibull. Por lo tanto, sólo si

interesara hacer inferencias con respecto a $Q_{0.95}$, podría utilizarse cualquiera de los tres modelos, y por sencillez se puede escoger un modelo Gumbel. Pero si el interés reside en obtener inferencias acerca del cuantil $Q_{0.99}$, es más conservador considerar la DGVE; esta es la práctica usual en aplicaciones de la Teoría de Valores Extremos.

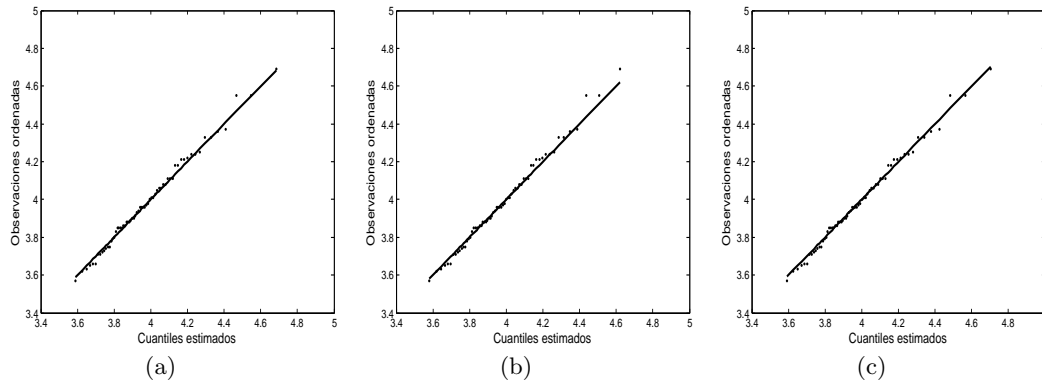


Figura 4.20: Datos de Port Pirie: Gráficas cuantil-cuantil usando los modelos (a) Gumbel (b) Weibull y (c) DGVE.

La Figura 4.20 muestra las gráficas cuantil-cuantil para los modelos Gumbel, Weibull y DGVE. Todos parecen ajustar bien a los datos, pero, debido a que se desconoce el tipo de distribución madre de las observaciones de donde se extrajeron los máximos, se suele preferir adoptar a la DGVE, lo cual es una posición conservadora. Por ello, considerando el análisis de la Figura 4.19, la sugerencia sería utilizar a la DGVE para realizar inferencias para cualquier cuantil.

En todas las gráficas anteriores, las verosimilitudes restringida y discretizada se traslaparon por completo y no hubo diferencia entre ellas. Nuevamente, ambas verosimilitudes perfiles permitieron fácilmente evaluar la robusticidad de las inferencias de parámetros de interés frente a modelos distintos, razonables para los datos.

4.3.2 Datos de voltajes de rompimiento

Hirose y Lai (1997), analizaron 20 mediciones de los voltajes de rompimiento de un fluido. Se contaba con un volumen de 4000 litros de aceite de transformador y se deseaba evaluar la resistencia que éste tenía como aislante frente al paso de corriente eléctrica. En este experimento interesa estimar un parámetro umbral, de manera que pueda afirmarse que no ocurrirán mediciones por debajo de ese umbral. Para modelar estos datos, los autores utilizaron una distribución Weibull de tres parámetros y una función de verosimilitud agrupada. Ellos muestran que para el modelo Weibull, las estimaciones puntuales que se obtienen con el método de Newton-Raphson, son inestables y en muchas ocasiones el emv toma valores infinitos. Por otra parte, también encuentran que aún utilizando una verosimilitud agrupada, se tienen los

mismos problemas de estimación. Su sugerencia es utilizar la DGVE debido a que el modelo Weibull es un caso particular de esta distribución y se tiene la ventaja que el estimador de máxima verosimilitud en la DGVE, es siempre un punto interior del espacio parametral. Su argumento principal es que con esta distribución, las estimaciones puntuales que se obtienen son siempre finitas, además de tener parámetros estables que simetrizan a la verosimilitud correspondiente.

Considerando las características de este conjunto de datos, el cual se presenta en la Sección C.5, del Apéndice C, un valor razonable para la precisión del instrumento de medición es $2h = 0.1$ volts. La media y varianza muestrales son $\bar{x} = 3.16$ y $s^2 = 0.1657$, respectivamente. La observación mínima es $x_{(1)} = 2.1$ volts y la máxima $x_{(20)} = 3.7$.

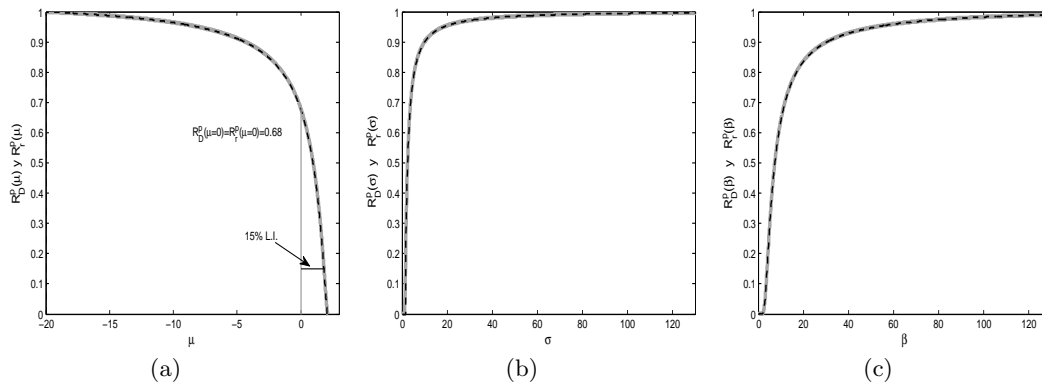


Figura 4.21: Datos de voltajes de rompimiento: Verosimilitud perfil restringida R_r^p (---) y discretizada R_D^p (—) de los parámetros (a) μ , (b) σ y (c) β de un modelo Weibull.

Como aún bajo el enfoque de una verosimilitud agrupada, Hirose y Lai encuentran un máximo local, alejado de la singularidad, y una verosimilitud perfil de μ plana y asimétrica, que favorece valores grandes y negativos para el parámetro umbral, sugieren estimar un cuantil por medio de un intervalo de confianza. Considerando esa sugerencia, aquí se estimará el cuantil de interés, pero por medio de su función de verosimilitud perfil y bajo los modelos Gumbel, Weibull y DGVE.

La verosimilitud perfil $R_r^p(\mu)$ del parámetro umbral μ es muy aplanada con la verosimilitud restringida, al igual que la verosimilitud discretizada perfil $R_D^p(\mu)$. Ambas conducen a resultados similares a los de Hirose y Lai, como puede observarse en la Figura 4.21(a). Esta gráfica resulta muy informativa porque indica que un modelo límite Gumbel, más sencillo, podría ser razonable para los datos. Por otra parte, como el valor $\mu = 0$ es plausible, ($R_D^p(\mu = 0) = 0.68$), incluso un modelo Weibull de dos parámetros, que es un modelo regular, podría ser razonable para este conjunto de datos.

Contrario a estas consideraciones, Smith and Naylor (1987b) y Green et al. (1994) critican una verosimilitud perfil plana del umbral $R_C^p(\mu)$ por considerarla ‘no informativa’, pasando por alto que contiene mucha información acerca de un modelo límite Gumbel. La sugerencia de considerar un modelo Gumbel, para analizar estos datos, también se deriva de la forma plana que toman las verosimilitudes perfiles de los parámetros σ y β del modelo Weibull analizado. Estas se construyeron tanto con la verosimilitud restringida como con la discretizada y coinciden en todas partes, como puede verse en las Figuras 4.21(b) y 4.21(c).

Cuando se analizan estos datos usando la DGVE se obtienen las estimaciones $\hat{a} = -3.364$, $\hat{b} = 0.27$ y $\hat{c} = 0.16$, tanto con la verosimilitud restringida como con la discretizada. La perfil del parámetro de forma c de esta distribución se muestra en la Figura 4.22, la cual se construyó utilizando ambas verosimilitudes. Puede observarse que el valor de $c = 0$ tiene una verosimilitud relativa alta, igual a 0.77. Esto indica que los tres modelos Weibull, Gumbel y Fréchet son razonables para analizar estos datos. Para fines de esta comparación se consideran solamente los modelos Weibull, Gumbel y DGVE, pues con el modelo Fréchet se obtienen resultados idénticos a los de la DGVE.

Aunque Hirose y Lai consideran el modelo Gumbel como una posible alternativa, lo descartan pues “*asigna una probabilidad positiva a valores negativos y parece ser incompatible con mediciones de voltajes, que siempre son positivas*”. Sin embargo con estos datos, los emv de los parámetros de localización y escala de un modelo Gumbel son 3.34 y 0.29 y bajo este modelo Gumbel, la probabilidad asignada a valores negativos es muy pequeña, $P[X < 0] = 0.000012$. Entonces, lo expuesto por estos autores, no es un argumento válido para descartar este modelo. Por otra parte, bajo el modelo Weibull estimado, de tres parámetros, la probabilidad de observar valores negativos es similar, ya que es aproximadamente 0.000015.

En muchos problemas un parámetro umbral μ es de interés por sí mismo, pero hay ocasiones, como en este ejemplo, que un modelo más simple como el Gumbel resulta apropiado y es ahí donde un cuantil pequeño puede proporcionar información equivalente al de una cota inferior.

Las simulaciones efectuadas en el Capítulo 3 mostraron que en ocasiones es difícil discernir entre un modelo Weibull y su modelo límite Gumbel, y esto ocurre más frecuentemente si el tamaño de la muestra es pequeño. Sin embargo, como los cuantiles demostraron cierta estabilidad, aquí se efectuará una comparación de los posibles modelos adecuados a los datos, en términos de las verosimilitudes perfiles de ciertos cuantiles.

En la Figura 4.23(a), se presenta la verosimilitud perfil del cuantil $Q_{.05}$ para el modelo Gumbel y los modelos DGVE y Weibull de tres parámetros. Estos mis-

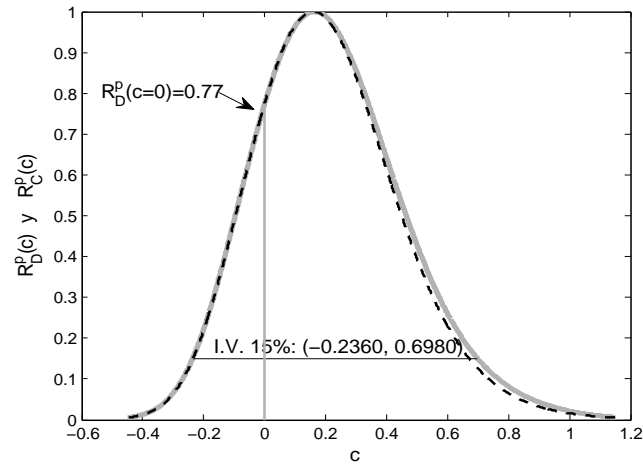


Figura 4.22: Datos de voltajes de rompimiento; Verosimilitud relativa perfil del parámetro de forma c de la DGVE, $R_r^p(c)$, (---) y $R_D^p(c)$ (—).

mos modelos se consideran para calcular las verosimilitudes perfil del cuantil $Q_{.50}$, mostradas en la Figura 4.23(b). Las perfiles de los cuantiles resultan muy informativas en este ejemplo, como puede verse en las Figuras 4.23(a) y 4.23(b), donde se aprecia que el modelo Weibull de tres parámetros es indistinguible del modelo Gumbel. Sin embargo, para el caso de $Q_{0.05}$, la verosimilitud perfil correspondiente al modelo de la DGVE, considera como plausible, valores más pequeños de este cuantil. Los intervalos del 15% de verosimilitud tanto para $Q_{0.05}$, como para $Q_{0.50}$, bajo estos diferentes modelos, se muestran en la Tabla 4.3.

	<i>I. V. del 15% para $Q_{0.05}$</i>	<i>I. V. del 15% para $Q_{0.50}$</i>
Weibull	(1.99, 2.77)	(3.07, 3.41)
Gumbel	(1.98, 2.77)	(3.08, 3.35)
DGVE	(0.28, 2.76)	(3.07, 3.41)

Tabla 4.3: Datos de voltajes de rompimiento: Intervalos del 15% de verosimilitud bajo los modelos Weibull, Gumbel y DGVE.

La verosimilitud relativa de $\mu = 0$ es alta, 0.68, por lo que el modelo Weibull de dos parámetros también podría considerarse adecuado. Sin embargo, se prefirió al modelo Weibull de tres parámetros por argumentos similares a los que se dieron en el ejemplo de las ratas analizado por Pike (1966).

En la Figura 4.24 se pueden observar también las gráficas cuantil-cuantil bajo los modelos Weibull, Gumbel y la DGVE. Todas parecen ajustar bien a los datos, aunque la gráfica correspondiente al modelo de la DGVE, mostrado en la Figura

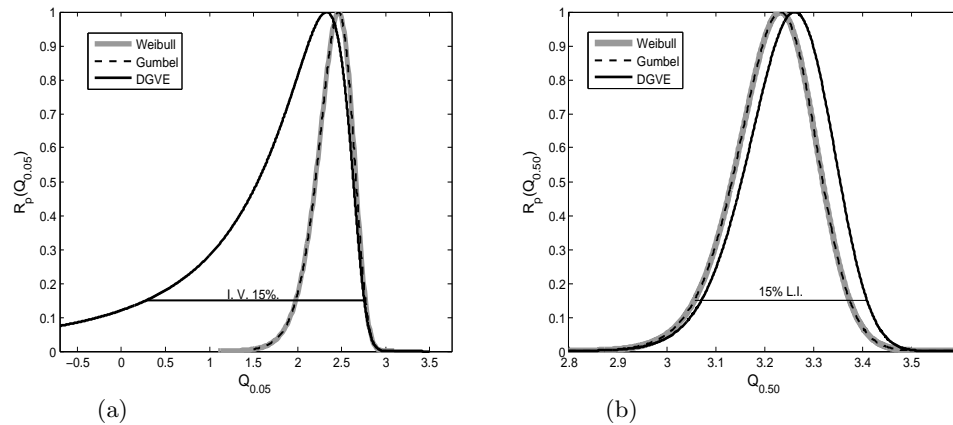


Figura 4.23: Datos de voltajes de rompimiento; Verosimilitud perfil bajo los modelos Weibull, Gumbel y DGVE, para los cuantiles (a) $Q_{0.05}$ y (b) $Q_{0.50}$.

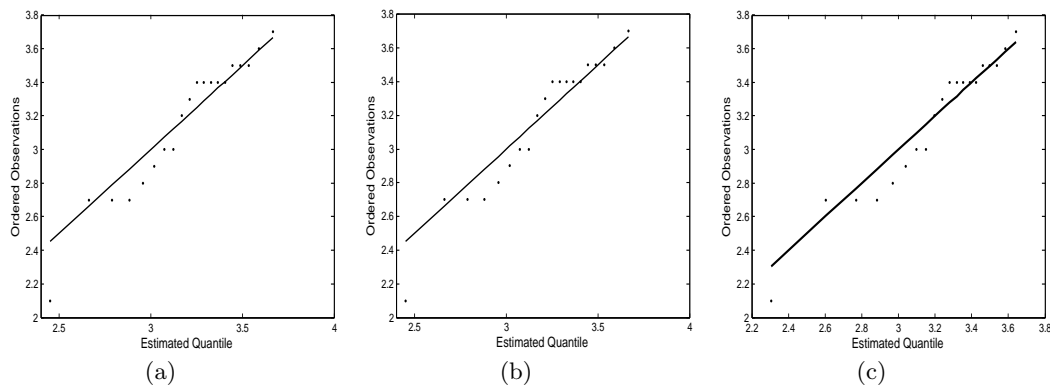


Figura 4.24: Datos de voltajes de rompimiento; Gráficas Q-Q bajo los modelos (a) Weibull de tres parámetros (b) Gumbel y (c) DGVE.

4.24(c), muestra un mejor ajuste para valores pequeños de los datos.

Después de analizar estos resultados se concluye que si el interés reside en estimar un cuantil central, como $Q_{0.50}$, puede utilizarse cualesquiera de estos modelos ya que se observa robusticidad bajo cambios en los modelos considerados y claramente, por sencillez, sería preferible un modelo Gumbel.

En contraste, las inferencias acerca de $Q_{0.05}$ no son robustas bajo los tres modelos. Como la muestra no es muy grande, una posición conservadora sería considerar al modelo DGVE como finalmente hacen Hirose y Lai (1997).

Aunque los intervalos de verosimilitud perfil del parámetro umbral tuvieron muy buenas coberturas, en las simulaciones realizadas para un escenario similar al pre-

sente, la estimación de los cuantiles demostró ser más robusta. Cuando se estimó el cuantil $Q_{0.05}$, bajo un modelo Gumbel las coberturas resultaron prácticamente las teóricas, desde muestras de tamaño 25. En cambio, cuando se utilizó un modelo Weibull, las coberturas teóricas se alcanzan a partir de muestras de tamaño 100. Todas estas consideraciones deben tomarse en cuenta en la selección del modelo más adecuado, pues en este caso se cuenta con tan sólo veinte observaciones.

4.3.3 Datos de dureza de fibras de vidrio

Smith y Naylor (1987a) utilizan la aproximación continua a la función de verosimilitud, para analizar un conjunto de $n = 46$ observaciones que obtienen cuando miden la fuerza necesaria para romper fibras de vidrio de 15 cms. de longitud, bajo condiciones experimentales. Estos experimentos son muy comunes en confiabilidad, y generalmente se busca establecer una cota inferior para estas mediciones de fuerza. Los autores consideraron una distribución Weibull, que es un modelo ampliamente utilizado para describir este tipo de comportamientos. Dado que en su análisis obtienen una estimación puntual negativa para el parámetro umbral y una verosimilitud perfil plana, argumentan que las técnicas de verosimilitud son numéricamente inestables. Su objetivo, entonces, es entender las razones de este comportamiento y mostrar métodos alternativos de estimación.

En este conjunto de datos, presentado en la Sección C.6 del Apéndice C, las mediciones de fuerza están reportadas con dos decimales, por lo que es razonable suponer una precisión de $2h = 0.01$, para el instrumento de medición. El dato mínimo es $x_{(1)} = 0.37$ y el máximo $x_{(46)} = 1.61$, mientras que la media y varianza muestrales son $\bar{x} = 1.13$ y $s^2 = 0.074$, respectivamente.

Aunque en realidad las mediciones de fuerza no puedan tomar valores negativos, la aproximación continua a la verosimilitud perfil asigna plausibilidades altas a valores negativos y grandes de μ , y su forma es asimétrica y plana. El mismo comportamiento resulta al utilizar la verosimilitud restringida y discretizada, como se observa en la Figura 4.25(a), donde puede verse que estas curvas se traslapan prácticamente en todas partes. Con ambas verosimilitudes, el intervalo al 15% de verosimilitud para el parámetro umbral μ resulta $(-\infty, 0.12)$.

Smith y Naylor argumentan que la función de densidad de la Weibull es inestable, en el sentido que pequeños cambios en la distribución originan cambios drásticos en los valores de los parámetros, esto debido a una pobre parametrización. Esta es una situación donde el modelo Weibull está cercano a su modelo Gumbel límite, por lo que modelos Weibull con parámetros aparentemente muy diferentes en magnitud describen a distribuciones muy cercanas. Esto se mostró en la Tabla 1.1, presentada en el Capítulo 1.

Las verosimilitudes perfil de los parámetros σ y β de este modelo Weibull, son también asimétricas y planas, como puede verse en las Figuras 4.25(b) y 4.25(c).

Como se mencionó en el ejemplo anterior, este resultado es muy valioso, pues la forma obtenida para estas verosimilitudes perfil, es otra vez un indicio de que un modelo Gumbel podría ser una buena alternativa para modelar estos datos.

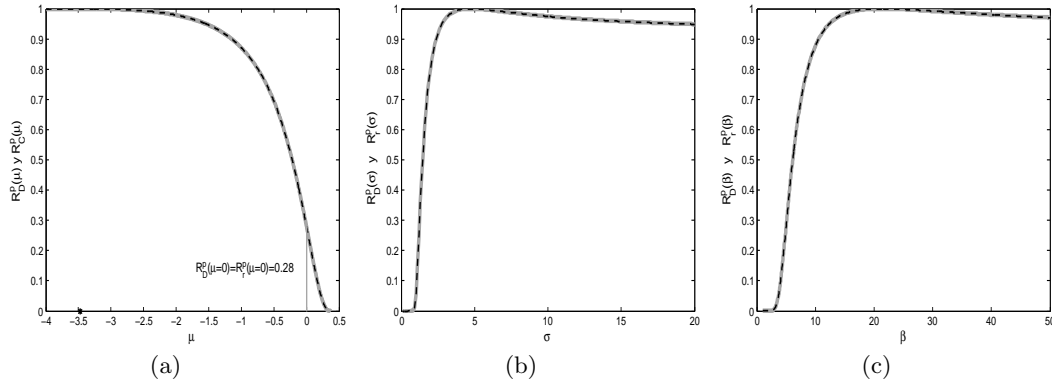


Figura 4.25: Datos de fuerza de fibra de vidrio: Verosimilitud perfil restringida R_r^p (---) y discretizada R_D^p (—) de los parámetros (a) μ , (b) σ y (c) β de un modelo Weibull.

Aunque Smith y Naylor ven como una posibilidad el trabajar con alguna reparametrización como son los cuantiles o bien la DGVE, no exploran éstas ya que su objetivo es mostrar que, en este problema, el enfoque Bayesiano proporciona estimaciones ‘razonables’. Ahora, con respecto a reparametrizar en términos de un cuantil, ellos tan solo estiman puntualmente el cuantil $Q_{0.01}$.

Cuando se analizan estos datos con la DGVE, bajo las verosimilitudes discretizada y restringida, se obtienen las estimaciones $\hat{a} = -1.2486$, $\hat{b} = 0.2215$ y $\hat{c} = -0.0469$. La Figura 4.26 muestra que las verosimilitudes relativas perfil restringida $R_r^p(c)$ y discretizada $R_D^p(c)$ del parámetro de forma c de la DGVE, se traslapan completamente. El intervalo del 15% de verosimilitud para este parámetro, que resulta $(-0.23, 0.19)$, contiene tanto valores positivos como negativos, por lo que cualquiera de los tres modelos Weibull, Gumbel y Fréchet son razonables. Como, $R_D^p(c=0) = 0.91$ sugiere un modelo Gumbel, mientras que el emv \hat{c} negativo, sugiere un modelo Weibull. Ambos modelos se exploran a continuación. El modelo Fréchet no mostró un buen ajuste a los datos y por ello no se incluye en este análisis.

Al igual que en el problema anterior, como principalmente interesa estimar una cota inferior para estos datos, se estimará un cuantil pequeño, como una estimación alternativa al parámetro umbral. Los modelos que parecen ser razonables a los datos se compararán por medio de las verosimilitudes perfiles del cuantil seleccionado y se evaluará su robusticidad.

Así pues, se analizarán, los modelos Weibull, Gumbel y la DGVE. Las Figuras

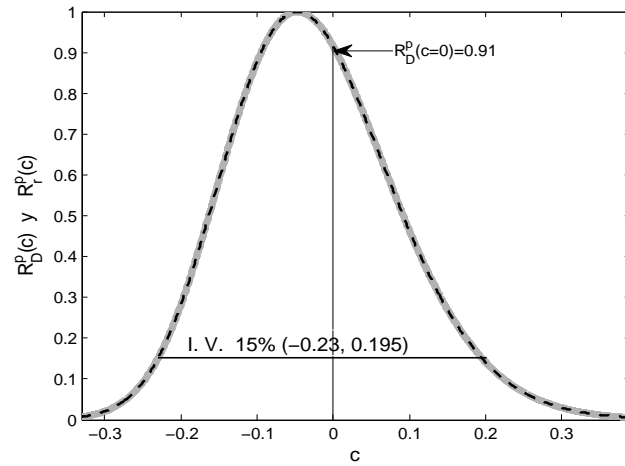


Figura 4.26: Datos de fuerza de fibras de vidrio; Verosimilitud perfil relativa del parámetro de forma c de la DGVE, $R_r^p(c)$, (---) y $R_D^p(c)$ (—).

4.27(a) y 4.27(b) muestran las verosimilitudes perfil de los cuantiles $Q_{0.05}$ and $Q_{0.50}$ bajo estos tres modelos. Si el objetivo fuese estimar la mediana de la distribución, entonces se podría seleccionar al modelo Gumbel. Sin embargo, si se desea estimar algún otro cuantil, es preferible seleccionar a la DGVE.

	<i>I. V. del 15% para $Q_{0.05}$</i>	<i>I. V. del 15% para $Q_{0.50}$</i>
Gumbel	(0.399, 0.759)	(1.090, 1.240)
Weibull	(0.499, 0.759)	(1.086, 1.242)
DGVE	(0.306, 0.781)	(1.086, 1.242)

Tabla 4.4: Datos de fibras de vidrio: Intervalos del 15% de verosimilitud bajo los modelos Weibull, Gumbel y DGVE.

La Tabla 4.4 contiene los intervalos del 15% de verosimilitud para $Q_{0.05}$ y $Q_{0.50}$, bajo cada uno de los modelos considerados. Considerando esta tabla y las verosimilitudes perfiles mostradas en la Figura 4.27, se observa que para $Q_{0.50}$, todos los modelos conducen a inferencias similares y se puede seleccionar el modelo más sencillo. Con respecto a las inferencias acerca del cuantil $Q_{0.05}$, los modelos Weibull y Gumbel son indistinguibles, mientras que los intervalos de verosimilitud obtenidos con la DGVE, son un poco más anchos. Ahora, el modelo Gumbel, que es más sencillo, asigna una probabilidad de 0.0032 a valores negativos, muy similar a la del modelo Weibull, que es igual a 0.0014; la DGVE tiene también una probabilidad asociada pequeña.

En la Figura 4.28 pueden observarse gráficas cuantil-cuantil para los modelos considerados. Aunque todas muestran un buen ajuste a los datos, la selección de la

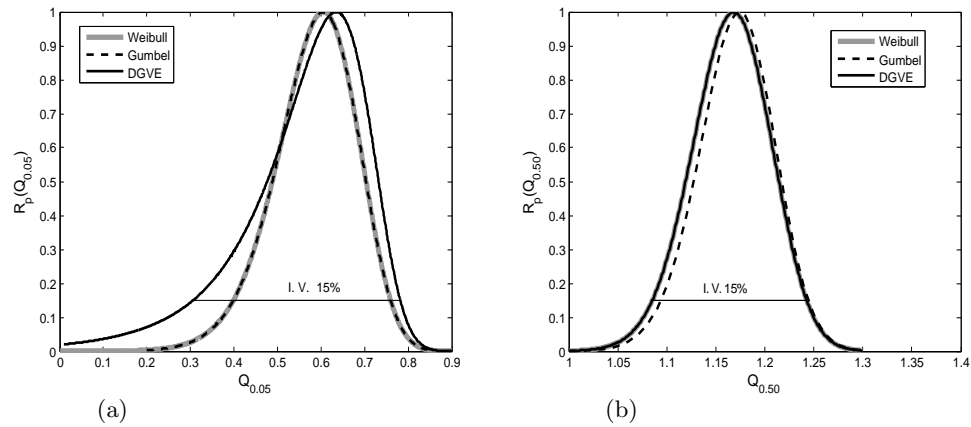


Figura 4.27: Datos de fuerza de fibras de vidrio; Verosimilitudes perfil relativa de los cuantiles (a) $Q_{0.05}$ y (b) $Q_{0.50}$ bajo los modelos Weibull, Gumbel y DGVE.

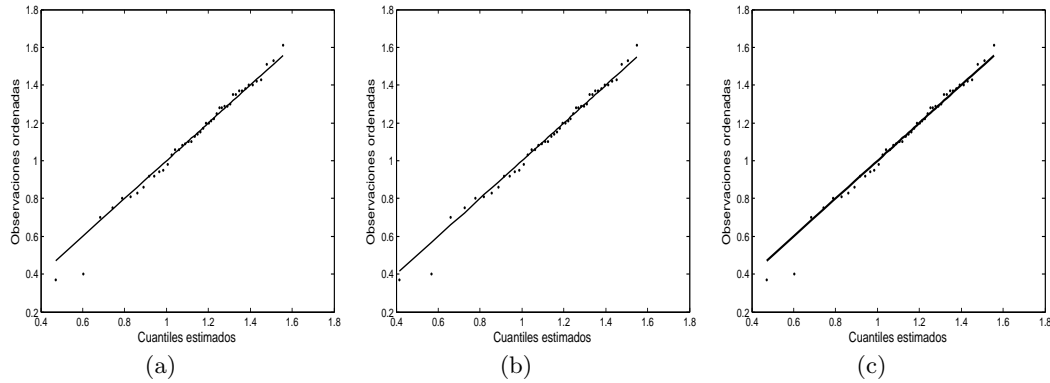


Figura 4.28: Datos de fuerza de fibra de vidrio: Gráficas Q-Q bajos los modelos (a) Weibull de tres parámetros (b) Gumbel y (c) DGVE.

DGVE es más conservadora con respecto a las inferencias, hecho que también puede observarse en la Figura 4.27(a).

Como puede verse, el elegir el modelo más adecuado a los datos, no es una tarea sencilla. Sin embargo, la función de verosimilitud perfil permitió evaluar diferentes aspectos que sirvieron para realizar un análisis más profundo y serio. Las simulaciones que se efectuaron emulando este escenario permiten proporcionar un valor agregado a la decisión que se tome, pues en este problema se cuenta con 46 observaciones y se observó que en muestras de tamaño 50, los intervalos de verosimilitud perfil de un cuantil pequeño como $Q_{0.05}$, tienen coberturas superiores a 0.92. Esto es importante pues en Teoría de Extremos suele recurrirse al uso de la DGVE directamente, sin evaluar la noción de robusticidad que se presentó no sólo en éste, sino en cada uno de los ejemplos analizados.

4.4 Discusión sobre ejemplos analizados

Todos los ejemplos analizados en este capítulo involucraron densidades con un parámetro umbral y singularidades. En todos ellos se confirma lo útil e informativo que resulta la función de verosimilitud perfil completa de los parámetros de interés. Sus gráficas permiten identificar fácilmente la asimetría de una superficie de verosimilitud, que al depender de varios parámetros no siempre es fácil de graficar. Por otra parte, la función de verosimilitud perfil permite identificar situaciones donde modelos más simples puedan ser razonables para describir al fenómeno aleatorio de interés. Ayuda también a evaluar la robusticidad de las inferencias sobre algún parámetro de interés, cuando se realizan cambios en el modelo estadístico seleccionado. Para ello, basta comparar las verosimilitudes perfil de este parámetro bajo cada uno de los modelos considerados. En aquellos casos donde las inferencias sobre los parámetros de interés sean robustas frente a cambios en el modelo, se sugiere considerar entonces, el modelo más sencillo.

En todos los ejemplos se observa que las funciones de verosimilitud perfil discretizada y restringida son prácticamente iguales en la mayoría de los casos, por lo que se pueden utilizar de manera indistinta. Ahora, con las facilidades actuales de cómputo, es igualmente sencillo utilizar cualquiera de estas dos verosimilitudes. Una ventaja de la verosimilitud discretizada es que permite considerar precisiones diferentes para cada una de las observaciones y sólo con ella es posible estimar de manera más precisa el parámetro umbral, en situaciones problemáticas donde la aproximación continua tiene forma de J . Este fue el único caso donde estas verosimilitudes discreparon un poco en cuanto al ancho de los intervalos de verosimilitud. Cabe señalar que una comparación de verosimilitudes como la aquí efectuada, nunca se había realizado en la literatura científica.

Para aquellos ejemplos donde el interés principal sea estimar un parámetro umbral, se mostró la conveniencia de utilizar parámetros alternativos como son los cuantiles, ya que suelen simetrizar la función de verosimilitud y son parámetros más estables. Una ventaja adicional es que el proceso de estimación numérica de los parámetros se vuelve mucho más eficiente.

En resumen, en todos los ejemplos analizados la verosimilitud discretizada y la verosimilitud restringida proporcionaron estimaciones razonables, en situaciones calificadas como problemáticas. Permitieron además, evaluar la robusticidad de las inferencias de los parámetros de interés, y contribuyeron en lograr una buena selección de modelos.

Capítulo 5

Conclusiones generales

El objetivo central en esta tesis fue el mostrar que la función de verosimilitud perfil discretizada resuelve los problemas de estimación que se presentan en ciertos modelos estadísticos, que se han considerado problemáticos en la literatura estadística, y que involucran densidades con parámetro umbral, singularidades y donde además se cuenta con muestras pequeñas. Se mostró también que la verosimilitud restringida propuesta por Barnard (1966) es una buena alternativa en estas situaciones. Esto se concluye gracias a las comparaciones que se efectuaron entre ellas y en las que resultaron ser similares. Sin embargo, se observa que la verosimilitud discretizada permite estimar con mayor precisión al parámetro umbral. El hecho que el verdadero umbral pueda tomar valores mayores que la observación más pequeña (en el caso de mínimos), o bien un valor menor que la observación más grande (en el caso de máximos) no se había considerado antes. Esta posibilidad puede ocurrir, sobre todo si la densidad teórica de donde proviene la muestra es convexa.

La verosimilitud restringida es analíticamente más sencilla que la verosimilitud discretizada; sin embargo, la verosimilitud discretizada incorpora de manera natural la precisión finita del instrumento de medición e incluso el hecho que las observaciones podrían tener precisión diferente. Con el poder computacional actual, es muy fácil de calcular, por lo cual se recomienda su uso para las situaciones antes descritas. La verosimilitud discretizada es similar a la de un modelo multinomial regular, por tanto posee las propiedades asintóticas de éste; esto es, se comporta como si fuese un caso regular. Este hecho se mostró claramente en las simulaciones realizadas en el Capítulo 3.

Las verosimilitudes perfiles, discretizada y restringida, se compararon a través de simulaciones, bajo diferentes escenarios de modelos Weibull de tres parámetros, para casos de muestras de tamaño moderado y pequeño, así como de diferentes precisiones del instrumento de medición. Las coberturas de los intervalos de verosimilitud perfil de los parámetros, que se obtuvieron con ambas verosimilitudes, mostraron por lo general ser buenas, incluso en muestras moderadas y pequeñas. Para los parámetros de escala y forma, con ambas verosimilitudes se observaron siempre buenas coberturas, aún en muestras pequeñas. En cambio, cuando la densidad teórica es convexa, el parámetro umbral requiere de muestras más grandes para poder lograr una buena estimación.

Aunque estimar un parámetro umbral puede ser de interés por la interpretación que éste tiene como una cota inferior (o superior) para la variable aleatoria en estudio, muchas veces es más conveniente estimar algún cuantil pequeño o grande. La reparametrización, en término de cuantiles, demostró ser una herramienta de gran utilidad. En las simulaciones realizadas en esta tesis, las coberturas de los intervalos de verosimilitud perfil para los cuantiles resultaron, en muchos de los escenarios analizados, incluso mejores que las que tuvieron los intervalos de estimación del parámetro umbral. Esto es, para un tamaño de muestra fijo, las coberturas de intervalos de verosimilitud perfil de un cuantil pequeño suelen ser mucho mejores que las de un parámetro umbral, por lo que resultan ser una parametrización más eficiente del modelo estadístico.

Algo importante de mencionar es que el máximo global de la verosimilitud discretizada y de la restringida, suelen coincidir con el máximo local de la aproximación continua a la verosimilitud, cuando éste existe. Esto justifica el uso generalizado que de este máximo local se ha hecho, considerándolo como el emv, en casos de densidades con singularidades.

En esta tesis, se mostró la importancia de analizar la forma de la función de verosimilitud perfil completa del parámetro umbral, ya que es muy informativa. Este es un aspecto al que no se le ha dado la debida importancia en la literatura estadística. Por ejemplo, cuando es aplanada en una región importante del espacio parametral, puede indicar que se trata de una situación donde la familia de modelos considerada está cercana a un modelo límite más simple, que podría ser más apropiado para describir al fenómeno aleatorio de interés.

Otro aspecto importante que se estudia en esta tesis es la estabilidad de los cuantiles y lo informativos que pueden ser éstos. Por lo general, los cuantiles suelen simetrizar la función de verosimilitud y hacer más eficiente el proceso de estimación. Esto es relevante pues cuando se logra una verosimilitud simétrica, las propiedades asintóticas de máxima verosimilitud se cumplen en muestras moderadas y en ocasiones incluso pequeñas. Esto facilita el realizar inferencias sobre los parámetros de interés, en términos de intervalos de verosimilitud-confianza. Por otra parte, el analizar la gráfica de la verosimilitud perfil de algún cuantil de interés permite evaluar la robusticidad de las inferencias frente a cambios en los modelos.

En resumen, se ha mostrado aquí que la verosimilitud perfil discretizada propuesta es una herramienta sencilla, informativa y poderosa que resuelve problemas de estimación importantes que surgen al usar métodos tradicionales para modelos no regulares, cuyas densidades tienen parámetro umbral y singularidades. Por tanto, se recomienda utilizarla en la inferencia científica.

Apéndice A

Resultados importantes de Estadística Matemática

A.1 Condiciones usuales de Regularidad

El teorema de máxima verosimilitud que se presenta en la Sección A.2, garantiza que los estimadores de máxima verosimilitud serán fuertemente consistentes, asintóticamente normales y asintóticamente eficientes, siempre que se satisfagan ciertas condiciones de regularidad en la familia $\mathfrak{F} = \{F_\theta, \theta \in \Theta\}$, donde $\Theta \subset \mathbb{R}^k$ y las distribuciones F_θ poseen densidades $f(x; \theta)$. Estas condiciones se retoman de (Serfling, 2002) y se enuncian a continuación.

Condiciones de Regularidad en \mathfrak{F} . Sea Θ un intervalo abierto, no necesariamente finito, en \mathbb{R} . Se supone

R1 Para $\theta \in \Theta$, las derivadas

$$\frac{\partial \log f(x; \theta)}{\partial \theta}, \quad \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2}, \quad \frac{\partial^3 \log f(x; \theta)}{\partial \theta^3}. \quad (\text{A.1.1})$$

existen.

R2 Para cada $\theta_0 \in \Theta$ existen funciones $g(x)$, $h(x)$ y $H(x)$ (posiblemente dependiente de θ_0), tales que, para θ en una vecindad $N(\theta_0)$ se cumplen las siguientes relaciones:

$$\left| \frac{\partial f(x; \theta)}{\partial \theta} \right| \leq g(x), \quad \left| \frac{\partial^2 f(x; \theta)}{\partial \theta^2} \right| \leq h(x), \quad \left| \frac{\partial^3 f(x; \theta)}{\partial \theta^3} \right| \leq H(x), \quad (\text{A.1.2})$$

donde $\int g(x)dx < \infty$, $\int h(x)dx < \infty$ y $E_\theta\{H(x)\} < \infty$ para $\theta \in N(\theta_0)$.

R3 Para cada $\theta \in \Theta$,

$$0 < E_\theta \left\{ \left(\frac{\partial \log f(x; \theta)}{\partial \theta} \right)^2 \right\} < \infty \quad (\text{A.1.3})$$

Así, ya descritas estas condiciones de regularidad, se enuncia enseguida el teorema de máxima verosimilitud, cuya demostración puede consultarse en Serfling (2002, pp. 145-148).

A.2 Teorema de Máxima Verosimilitud

La demostración del siguiente teorema puede consultarse en Serfling (2002, p. 153)

Teorema A.1 *Supóngase se satisfacen las condiciones de regularidad **R1**, **R2** y **R3** en la familia \mathfrak{F} y considere observaciones i.i.d. en F_θ , para $\theta \in \Theta$. Entonces, con probabilidad 1, las ecuaciones de verosimilitud admiten una sucesión de soluciones $\{\hat{\theta}_n\}$ que satisfacen*

(i) *Consistencia fuerte:* $\hat{\theta} \xrightarrow{p} \theta, n \rightarrow \infty$

(ii) *Normalidad asintótica y eficiencia:*

$$w(\theta) = (\hat{\theta}_n - \theta) \sqrt{n\mathcal{I}(\theta)} \xrightarrow{d} N(0, 1),$$

donde $\mathcal{I}(\theta) = -E_\theta (\partial^2 l(\theta; x_i) / \partial \theta^2)$ es la información esperada de una sola observación.

Este teorema se puede generalizar para el caso multiparamétrico donde $\theta = (\theta_1, \dots, \theta_k)$, ya que bajo una apropiada generalización de las condiciones de regularidad **R1**, **R2** y **R3**, descritas en Serfling (2002, p. 148), existe una sucesión $\hat{\theta}_n$ de soluciones a las ecuaciones de máxima verosimilitud, tales que

(i) $\hat{\theta}_n \xrightarrow{p} \theta$ y

(ii) $(\hat{\theta} - \theta) \xrightarrow{d} N(0, n^{-1}\mathbf{I}_\theta^{-1})$,

donde \mathbf{I}_θ es la matriz de información dada por

$$\mathbf{I}_\theta = \left[E_\theta \left\{ \frac{\partial \log f(X; \theta)}{\partial \theta_i} \cdot \frac{\partial \log f(X; \theta)}{\partial \theta_j} \right\} \right]_{k \times k}.$$

A.3 Distribución asintótica del estadístico de la razón de verosimilitudes

La demostración del siguiente teorema puede consultarse en Serfling (2002, p. 155) y en Pawitan (2007, p. 266).

Teorema A.2 *Sean X_1, X_2, \dots variables aleatorias independientes con distribución F_θ , $\theta \in \mathbb{R}^k$. Bajo condiciones apropiadas de regularidad, la cantidad pivotal*

$$\Lambda_n = -2 \log \frac{L(\theta; \mathbf{y})}{L(\hat{\theta}; \mathbf{y})}$$

converge en distribución a una χ_k^2 .

El estadístico de la razón de verosimilitudes puede también calcularse utilizando la verosimilitud perfil. Esto es, si $\theta = (\theta_1, \theta_2) \in \mathbb{R}^k$, donde $\theta_1 \in \mathbb{R}^q$ es un parámetro q -dimensional de interés y $\theta_2 \in \mathbb{R}^s$, es el parámetro s -dimensional de estorbo, con $k = q + s$, entonces Λ_n seguirá una distribución Ji-cuadrada con $k - s$ grados de libertad.

Apéndice B

Algunos cálculos relevantes

B.1 Modelo límite para la distribución Weibull

A continuación se muestra cómo el modelo Gumbel, con parámetro de localización θ y ψ un parámetro de escala, es un modelo límite, de dos parámetros, para el modelo Weibull con parámetro umbral μ y parámetros de escala y forma σ y β , respectivamente.

Si en la función de densidad del modelo Weibull

$$f(y; \mu, \sigma, \beta) = \frac{\beta}{\sigma} \left(\frac{y - \mu}{\sigma} \right)^{\beta-1} \exp \left[- \left(\frac{y - \mu}{\sigma} \right)^\beta \right] I_{[\mu, \infty)}(y), \quad (\text{B.1.1})$$

se utiliza la reparametrización $\sigma = \psi\beta$ y $\mu = \theta - \psi\beta$, entonces

$$f(y; \theta, \psi, \beta) = \frac{\beta}{(\psi\beta)^\beta} (y - \theta + \psi\beta)^{\beta-1} \exp \left\{ - \left(\frac{y - \theta + \psi\beta}{\psi\beta} \right)^\beta \right\} \quad (\text{B.1.2})$$

$$= \frac{1}{\psi} \left(\frac{y - \theta}{\psi\beta} + 1 \right)^{\beta-1} \exp \left\{ - \left(\frac{y - \theta}{\psi\beta} + 1 \right)^\beta \right\}, \quad (\text{B.1.3})$$

de donde, tomando el $\lim_{\beta \rightarrow \infty}$, se obtiene

$$\lim_{\beta \rightarrow \infty} \frac{1}{\psi} \left(\frac{y - \theta}{\psi\beta} + 1 \right)^{\beta-1} \exp \left\{ - \left(\frac{y - \theta}{\psi\beta} + 1 \right)^\beta \right\} = \frac{1}{\psi} \exp \left\{ \frac{y - \theta}{\psi} - \exp \frac{y - \theta}{\psi} \right\}, \quad (\text{B.1.4})$$

que corresponde a la función de densidad de un modelo Gumbel, donde $\psi = \sigma/\beta$ y $\theta = \mu + \psi\beta$.

B.2 La verosimilitud discretizada para el caso exponencial con t.v.g.

En esta sección se presentan algunos cálculos que se omitieron en la Sección 3.5.2, donde se analiza la distribución del estadístico de la razón de verosimilitudes, cuando se utiliza la verosimilitud discretizada L_D , en la distribución exponencial con tiempo

de vida garantizado.

La verosimilitud discretizada se expresó como:

$$L_D(\mu, \sigma; \mathbf{x}) = L_1(\mu, \sigma; \mathbf{x}) + L_2(\mu, \sigma; \mathbf{x}) \quad (\text{B.2.1})$$

donde

$$L_1(\mu, \sigma; \mathbf{x}) = \exp\left(-\sum_{i=1}^n \frac{x(i)}{\sigma}\right) \times \left[\exp\left(\frac{\mu+h}{\sigma}\right) - \exp\left(\frac{\mu-h}{\sigma}\right)\right]^n I_{(-\infty, x(1)-h]}(\mu) I_{(0, \infty)}(\sigma) \quad (\text{B.2.2})$$

y

$$L_2(\mu, \sigma; \mathbf{x}) = \left[1 - \exp\left(-\frac{x(1)+h-\mu}{\sigma}\right)\right]^m \exp\left(-\sum_{i=m+1}^n \frac{x(i)}{\sigma}\right) \times \left[\exp\left(\frac{\mu+h}{\sigma}\right) - \exp\left(\frac{\mu-h}{\sigma}\right)\right]^{n-m} I_{(x(1)-h, x(1)+h]}(\mu) I_{(0, \infty)}(\sigma). \quad (\text{B.2.3})$$

La muestra discretizada $\mathbf{x} = x(1), \dots, x(n)$, se considera medida con un instrumento de medición que posee una precisión fija $2h$.

A continuación se analizan sólo algunos de los cuatro escenarios posibles de la Tabla 3.1, presentada en la Sección 3.5.2.

B.2.1 Distribución de $-2 \log R_D(\theta_0; \mathbf{x})$ en muestras de tamaño uno y σ conocida

Como se explicó en la Sección 3.5.2, cuando se analizan muestras de tamaño uno y σ conocida, se obtiene que $\hat{\mu} = x(1) - h$ y solamente se analiza el caso que a continuación se presenta.

Caso 1. En este Caso, tanto $\hat{\mu}$ como μ_0 están dentro del intervalo $(-\infty, x(1) - h]$ y como

$$\begin{aligned} L_D(\mu, \sigma = \sigma_0; \mathbf{x}) &= L_1(\mu, \sigma = \sigma_0; \mathbf{x}) = P(x(1) - h < X < x(1) + h) \\ &= \exp\left[-\frac{(x(1) - h - \mu)}{\sigma}\right] - \exp\left[-\frac{(x(1) + h - \mu)}{\sigma}\right] \quad \text{por (3.5.1)} \\ &= \exp\left(-\frac{x(1)}{\sigma}\right) \exp\left(\frac{\mu}{\sigma}\right) \left[\exp\left(\frac{h}{\sigma}\right) - \exp\left(-\frac{h}{\sigma}\right)\right] = k_1 \exp\left(\frac{\mu}{\sigma}\right), \end{aligned}$$

donde $k_1 > 0$ y el emv se alcanza siempre en el extremo derecho del soporte de $L_1(\mu, \sigma; \mathbf{x})$, por lo que se tiene que $\hat{\mu} = x(1) - h$.

El comportamiento de $-2 \log R(\mu_0; \mathbf{x})$ para este caso es,

$$-2 \log R_D(\mu_0; \mathbf{x}) = -2 \log \left[\frac{L_1(\mu_0, \sigma = \sigma_0; \mathbf{x})}{L_1(\hat{\mu}, \sigma = \sigma_0; \mathbf{x})} \right], \quad (\text{B.2.4})$$

y

$$\begin{aligned} \frac{L_1(\mu_0, \sigma = \sigma_0; \mathbf{x})}{L_1(\hat{\mu}, \sigma = \sigma_0; \mathbf{x})} &= \frac{\exp(-x_{(1)}/\sigma_0) \cdot \exp(\mu_0/\sigma_0) \left[\exp(h/\sigma_0) - \exp(-h/\sigma_0) \right]}{\exp(-x_{(1)}/\sigma_0) \cdot \exp(\hat{\mu}/\sigma_0) \left[\exp(h/\sigma_0) - \exp(-h/\sigma_0) \right]} \\ &= \exp \left[-\frac{(x_{(1)} - h - \mu_0)}{\sigma_0} \right], \end{aligned}$$

por lo que,

$$-2 \log R_D(\mu_0; \mathbf{x}) = 2 \left(\frac{x_{(1)} - h - \mu_0}{\sigma_0} \right).$$

Con lo anterior podemos calcular la distribución de $-2 \log R_D(\mu_0; \mathbf{x})$, esto es

$$\begin{aligned} P(-2 \log R_D(\mu_0; \mathbf{x}) \leq z) &= P \left[2 \left(\frac{x_{(1)} - h - \mu_0}{\sigma_0} \right) \leq z \right] \\ &= P \left(x_{(1)} \leq \mu_0 + \frac{z\sigma_0}{2} + h \right) \\ &= 1 - \exp \left[-(\mu_0 + \frac{z\sigma_0}{2} + h - \mu_0)/\sigma_0 \right] \\ &= 1 - \exp(-z/2) \cdot \exp(-h/\sigma_0) \end{aligned} \quad (\text{B.2.5})$$

Usando nuevamente el argumento visto en la Sección 3.5.1, si una variable aleatoria X sigue una distribución $\chi_{(2)}^2$, entonces la variable $Z = X - 2h/\sigma_0$ tiene una distribución dada por

$$\begin{aligned} P(Z \leq z) &= P \left(X - \frac{2h}{\sigma_0} \leq z \right) = P \left(X \leq z + \frac{2h}{\sigma_0} \right) \\ &= 1 - \exp(-z/2) \cdot \exp(-h/\sigma_0). \end{aligned} \quad (\text{B.2.6})$$

De donde se concluye que $-2 \log R(\mu_0; \mathbf{x})$ sigue una distribución $\chi_{(2)}^2$, trasladada una distancia $2h/\sigma_0$.

B.2.2 Distribución de $-2 \log R_D(\mu_0; \mathbf{x})$ en muestras de tamaño n y σ conocida

Para muestras de tamaño n y σ conocida se omitieron algunos cálculos del Caso 4 de la Tabla 3.1, presentada en la Sección 3.5.2, donde $\hat{\mu} \in (x_{(1)} - h, x_{(1)} + h]$ y $\mu_0 \in (x_{(1)} - h, x_{(1)} + h]$. En este caso

$$\begin{aligned} L_D(\mu, \sigma = \sigma_0; \mathbf{x}) &= L_2(\mu, \sigma = \sigma_0; \mathbf{x}) = \left\{ 1 - \exp \left[-\frac{(x_{(1)} + h - \mu)}{\sigma_0} \right] \right\}^m \times \\ &\exp \left(-\sum_{i=m+1}^n \frac{x_{(i)}}{\sigma_0} \right) \left[\exp \left(\frac{\mu + h}{\sigma_0} \right) - \exp \left(\frac{\mu - h}{\sigma_0} \right) \right]^{n-m} I_{(x_{(1)} - h, x_{(1)} + h]}(\mu). \end{aligned}$$

Calculando el logaritmo de L_D se tiene que

$$l_D(\mu, \sigma = \sigma_0; \mathbf{x}) = m \log \left\{ 1 - \exp \left[-\frac{(x_{(1)} + h - \mu)}{\sigma_0} \right] \right\} - \sum_{i=m+1}^n \frac{x_{(i)}}{\sigma_0} \\ + (n - m) \log \left[\exp \left(\frac{\mu + h}{\sigma_0} \right) - \exp \left(\frac{\mu - h}{\sigma_0} \right) \right], \quad (\text{B.2.7})$$

y suponiendo que L_2 es diferenciable, se calcula la función Score $S_D(\mu, \sigma = \sigma_0; \mathbf{x})$, a partir de la cual puede obtenerse el estimador de máxima verosimilitud de μ . Así,

$$S_D = \frac{m \exp [-(x_{(1)} + h)/\sigma_0]}{\exp(-\mu/\sigma_0) - \exp [-(x_{(1)} + h)/\sigma_0]} + n - m = 0, \quad (\text{B.2.8})$$

y el estimador máximo verosímil de μ es

$$\hat{\mu} = -\sigma_0 \log \left[\frac{n}{n - m} \cdot \exp \left(-\frac{x_{(1)} + h}{\sigma_0} \right) \right] \\ = (x_{(1)} + h) - \sigma_0 \log \left(\frac{n}{n - m} \right) \\ = (x_{(1)} + h) + \sigma_0 \log \left(1 - \frac{m}{n} \right) \quad (\text{B.2.9})$$

donde m es el número de veces que se repite la observación más pequeña.

El resultado anterior podríamos analizarlo de la manera siguiente. Consideremos que

$$p = P(\mu < X \leq x_{(1)} + h) \quad (\text{B.2.10})$$

esto es $p = 1 - \exp [(\mu - h - x_{(1)})/\sigma_0]$, de donde obtenemos que

$$\mu = (x_{(1)} + h) + \sigma_0 \log(1 - p). \quad (\text{B.2.11})$$

Así, podemos entonces ver que el emv de μ obtenido en (B.2.9), puede escribirse como

$$\hat{\mu} = (x_{(1)} + h) + \sigma_0 \log(1 - p_{m,n}), \quad (\text{B.2.12})$$

que sería la versión empírica de la expresión (B.2.11), donde $p_{m,n} = m/n$.

B.2.3 Distribución de $-2 \log R_D(\mu_0; \mathbf{x})$ en muestras de tamaño n y σ desconocida

Cuando el parámetro de escala σ , es desconocido, debe también estimarse, ya que sólo de esta manera se puede estudiar el comportamiento de $-2 \log R(\theta_0; \mathbf{x})$.

Para el Caso 1 y en muestras de tamaño n , la función de verosimilitud discretizada está dada por

$$L_D(\mu, \sigma; \mathbf{x}) = \prod_{i=1}^n \{\exp[-(x_i - h - \mu)/\sigma] - \exp[-(x_i + h - \mu)/\sigma]\}. \quad (\text{B.2.13})$$

Calculando su logaritmo

$$l_D(\mu, \sigma; \mathbf{x}) = -\frac{\sum_{i=1}^n x_i}{\sigma} + \frac{n\mu}{\sigma} + n \log [\exp(h/\sigma) - \exp(-h/\sigma)], \quad (\text{B.2.14})$$

derivando con respecto de σ y sustituyendo $\hat{\mu} = x_{(1)} - h$ como el estimador de máxima verosimilitud de μ , se llega a que el estimador de σ está dado por la expresión

$$\hat{\sigma} = 2h \cdot \left[\log \left(\frac{\bar{x} - \hat{\mu} + h}{\bar{x} - \hat{\mu} - h} \right) \right]^{-1}. \quad (\text{B.2.15})$$

Por otra parte, para el Caso 4, donde el emv de μ resulta

$$\begin{aligned} \hat{\mu} &= -\sigma \log \left(\frac{n}{n-m} \right) \cdot \exp \left[- \left(\frac{x_{(1)} + h}{\sigma} \right) \right] \\ &= (x_{(1)} + h) - \sigma \log \left(\frac{n}{n-m} \right), \end{aligned}$$

como se mostró en la Sección B.2.2, al derivar B.2.7 con respecto de σ y sustituir este e.m.v. de μ , se llega a que el estimador para el parámetro de escala σ también está dado por (B.2.15).

Apéndice C

Algunos conjuntos de datos analizados

A continuación se presentan los conjuntos de datos que se utilizaron en algunos de los ejemplos del Capítulo 4 de esta tesis.

C.1 Diámetros de árboles de cerezas negras

Los siguientes datos representan 31 mediciones de los diámetros de árboles adultos de cerezas negras (Ryan et al., 1985), medidos 54 pulgadas arriba del nivel del suelo.

8.3, 8.6, 8.8, 10.5, 10.7, 10.8, 11, 11, 11.1, 11.2, 11.3, 11.4, 11.4, 11.7, 12, 12.9, 12.9, 13.3, 13.7, 13.8, 14, 14.2, 14.5, 16, 16.3, 17.3, 17.5, 17.9, 18, 18, 20.6

C.2 Tiempos de detección de cáncer en ratas

Los siguientes se obtuvieron de un experimento de inducción de cáncer realizado por Pike (1966) en el que se pintó la vagina de 19 ratas de laboratorio con un cancerígeno y se procedió a contar el número T de días, que transcurrieron hasta detectar un carcinoma. Hay dos observaciones censuradas, denotadas con asterisco, para las cuales no hay indicios de carcinoma alguno hasta el día registrado.

143, 164, 188, 188, 190, 192, 206, 209, 213, 216, 220, 227, 230, 234, 246, 265, 304, 216*, 244*

C.3 Tiempos de espera de ratas para ejecutar tareas de aprendizaje

La información que se presenta a continuación, corresponde a los tiempos de latencia de retención en dos grupos de ratas, de un experimento que se llevó a cabo en el Instituto de Neurobiología de la UNAM, campus Querétaro, bajo la dirección del Dr. Roberto Prado- Alcalá.

- Grupo de ratas bajo el placebo: Dato mínimo: $x_{(1)} = 4.32$, dato mayor: $x_{(12)} = 78.9$, media: $\bar{x} = 6.08$ y desviación estándar: $s = 21.68$.

- Grupo de ratas bajo la droga: Dato mínimo: $x_{(1)} = 2.77$, dato mayor: $x_{(12)} = 65.74$, media: $\bar{x} = 31.15$ y desviación estándar: $s = 19.49$.

C.4 Máximos anuales del nivel del mar en Port Pirie, Australia

Los siguientes datos corresponden a los máximos anuales de las alturas diarias del nivel del mar, entre 1923 y 1987, en Port Pirie, Australia. Estos datos pueden obtenerse, del sitio

<http://www.maths.lth.se/matstat/kurser/fms155mas231/datasetsR.html>,

donde se encuentran capturados en una matriz de datos con nombre *portpirie.R*.

154, 49.6, 46.7, 58.3, 70.5, 90, 70.1, 105.7, 37.4, 40.8, 34.7, 58.9, 72.2, 30, 71.6, 100, 33.7, 49.9, 56.1, 142.3, 28.6, 54.8, 74.1, 60, 50.9, 38.6, 53.4, 132.5, 50.7, 40.8, 84.3, 38.8, 27.4, 67, 118.7, 23.2 55, 67.9, 87.3, 89, 98.7, 47.1, 71.6, 83.6, 44.3, 41.2, 35.9, 44.3, 410.4

C.5 Voltajes de rompimiento de fluidos

Los siguientes datos son presentados y analizados por Hirose y Lai (1997) y corresponden a 20 mediciones de los voltajes de rompimiento de un fluido.

2.1, 3.4, 3.7, 2.7, 3.2, 3.5, 2.7, 2.7, 3.4, 3.3, 3.0, 2.9, 3.0, 3.4, 2.8, 3.5, 3.5, 3.4, 3.4, 3.6

C.6 Dureza de fibras de vidrio

El siguiente conjunto de datos, analizado por Smith y Naylor (1987a), representan las mediciones de la fuerza necesaria para romper fibras de vidrio de 15 centímetros de longitud.

0.37, 0.81, 0.94, 1.06, 1.13, 1.20, 1.28, 1.35, 1.40, 1.61, 0.40, 0.83, 0.95, 1.08, 1.14, 1.21, 1.29, 1.37, 1.42, 0.70, 0.86, 0.98, 1.09, 1.15, 1.22, 1.29, 1.37, 1.43, 0.75, 0.92, 1.03, 1.10, 1.17, 1.25, 1.30, 1.38, 1.51, 0.80, 0.92, 1.06, 1.10, 1.20, 1.28, 1.35, 1.40, 1.53

Bibliografía

- Bailey, R. L. and Dell, T. R. (1973). Quantifying diameter distributions with the Weibull function. *Forest Science*, 19(2):97–104.
- Barnard, G. A. (1966). The use of the likelihood function in statistical practice. *5th Berkeley Symp.*, 1:27–40.
- Barndorff-Nielsen, O. E. and Cox, D. R. (2001). *Inference and Asymptotics*. Chapman & Hall, New York.
- Berger, J. O., Liseo, B., and Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(1):1–22.
- Bolívar, A. M. (2007). Intervalos de estimación bootstrap para cuantiles en teoría de valores extremos. Master's thesis, CIMAT.
- Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society*, 143(4):765–791.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B.*, 26(2):211–252.
- Cheng, R. C. H. and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(3):394–403.
- Cheng, R. C. H. and Iles, T. C. (1990). Embedded models in three-parameter distributions and their estimation. *Journal of the Royal Statistical Society. Series B.*, 52(1):135–149.
- Coles, S. (2004). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, third edition.
- Cox, D. R. and Hinkley, D. V. (2000). *Theoretical Statistics*. Chapman & Hall, New York.
- Díaz-Trujillo, A., C. J. M. A. S.-L. G. A. A. Q. G. P.-A. R. A. (2009). Enhanced inhibitory avoidance learning prevents the long-term memory-impairing effects of cycloheximide, a protein synthesis inhibitor. *Neurobiology of Learning and Memory*, (91):310–314.
- Edwards, A. W. F. (1992). *Likelihood*. The Johns Hopkins University Press, Baltimore and London, expanded edition.
- Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32.

- Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London, Series A.*, 144:285–307.
- Fisher, R. A. (1941). The negative binomial distribution. *Annals of Eugenics*, 6.
- Fisher, R. A. (1945). The logical inversion of the notion of the random variable. *Sankhya*, 7(2):129–132.
- Gentleman, R. (2004). Some perspectives on statistical computing. *The Canadian Journal of Statistics*, 32(3):209–226.
- Giesbrecht, F. and Kempthorne, O. (1976). Maximum likelihood estimation in the three-parameter lognormal distribution. *Journal of the Royal Statistical Society, Series B.*, 38(3):257–264.
- Gómez, G., Calle, L. M., and Oiler, R. (2004). Frequentist and Bayesian approaches for interval-censored data. *Statistical Papers*, 45:139–173.
- Green, E. J., Roesch, F. A. J., Smith, A. F. M., and Strawderman, W. E. (1994). Bayesian estimation for the three-parameter Weibull distribution with tree diameter data. *Biometrics*, 50(1):254–269.
- Heathcote, A., Brown, S., and Cousineau, D. (2004). QMPE: Estimating lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments, & Computers*, 36(2):277–290.
- Hill, B. M. (1963). The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *Journal of the American Statistical Association*, 58(301):72–84.
- Hirose, H. and Lai, T. L. (1997). Inference from grouped data in three parameter Weibull models with applications to breakdown-voltage experiments. *Technometrics*, 39(2):199–210.
- Hood, W. C. and Koopmans, T. C. E. (1953). *Studies in Econometric Method*. Wiley, New York.
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Quarterly J. Royal Meteorological Society*, (81):158–172.
- Kalbfleisch, J. D. and Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society, Series B.*, 32(2):175–208.
- Kalbfleisch, J. G. (1985). *Probability and Statistical Inference*, volume 2. Oxford University Press, New York, second edition.
- Kempthorne, O. (1966). Some aspects of experimental inference. *J. Amer. Statist. Ass.*, 61:11–34.

- Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. World Scientific Publishing Company, London, 1st edition.
- Lauro, C. (1996). Computational statistics or statistical computing, is that the question? *Computational Statistics & Data Analysis*, 23:191–193.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley-Interscience, New York, second edition.
- Lindsey, J. K. (1999). Some statistical heresies. *The Statistician*, 48(1):1–40.
- Mann, N. (1984). *Statistical estimation of parameters of the Weibull and Fréchet distributions*. In *Statistical Extremes and Applications* (ed Tiago de Oliveira J). Reidel Publishing Company, Holland.
- Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley Series in Probability and Statistics, New York.
- Mises, R. V. (1954). La distribution de la plus grande de n valeurs. *Selected Papers, V. II, American Mathematical Society*, pages 271–294.
- Montoya, J. A. (2008). *La verosimilitud perfil en la Inferencia Estadística*. PhD thesis, CIMAT.
- Montoya, J. A., Díaz-Francés, E., and Sprott, D. A. (2009). On a criticism of the profile likelihood function. *Statistical Papers*, 50:195–202.
- Odell, P. M., Anderson, K. M., and D’Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time. *Biometrics*, 48(3):951–959.
- Pawitan, Y. (2007). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press Inc., New York.
- Pike, M. C. (1966). A method of analysis of certain classes of experiments in carcinogenesis. *Biometrics*, 22:142–161.
- Ross, G. J. S. (1970). The efficient use of function minimization in non-linear maximum likelihood estimation. *Journal of the Royal Statistical Society. Series C.*, 19(3):205–221.
- Ryan, B. F., Joiner, B. L., and Ryan, T. A. J. (1985). *Minitab Handbook*. Duxbury Press, Boston: MA, second edition.
- Seber, G. A. F. and Wild, C. J. (2003). *Non Linear Regression*. John Wiley & Sons, New Jersey.
- Serfling, R. J. (2002). *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons, New York.
- Shibata, E., MacIsaac, D., and Weissman, A. (1995). *Purdue University Physics 152L Manual*. Lafayette, IN., fifth edition. 149p.

- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.
- Smith, R. L. and Naylor, J. C. (1987a). A comparison of maximum likelihood and Bayesian estimators for the three parameter Weibull distribution. *Journal of the Royal Statistical Society*, 36(3):358–369.
- Smith, R. L. and Naylor, J. C. (1987b). Statistics of the three-parameter Weibull distribution. *Annals of Operations Research*, 9:577–587.
- Smith, R. L. and Weissman, I. (1985). Maximum likelihood estimation of the lower tail of a probability distribution. *Journal of the Royal Statistical Society. Series B.*, 47(2):285–298.
- Sprott, D. A. (2000). *Statistical Inference in Science*. Springer, New York.
- Sprott, D. A. and Farewell, V. T. (1993). The difference between two normal means. *The American Statistician*, 47(2):126–128.
- Sprott, D. A. and Viveros, R. (1984). The interpretation of maximum-likelihood estimation. *The Canadian Journal of Statistics*, 12(1):27–38.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, (9):60–62.