



"El saber de mis hijos
hará mi grandeza"

UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

Programa de Licenciado en Matemáticas

Teoría de Valores de Extremos Aplicada a la Gestión de
Riesgos en Inundaciones

T E S I S

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

Gilberto Molina Aguilar

Director de Tesis: Dr. José Arturo Montoya Laos

Hermosillo, Sonora, México, 8 de Diciembre, 2010.

SINODALES

Dr. José Arturo Montoya Laos
Universidad de Sonora

M.C. Ma. Teresa Robles Alcaráz
Universidad de Sonora

M.C. Gudelia Figueroa Preciado
Universidad de Sonora

Dr. Oscar Vega Amaya
Universidad de Sonora

Agradecimientos

Al Dr. José Arturo Montoya Laos por dirigir este trabajo de tesis. Gracias maestro por su paciencia, su dedicación y sobre todo sus consejos para llevar acabo este trabajo. También, quiero agradecer al comité revisor de esta tesis la M.C. Ma. Teresa Robles Alcaráz, la M.C. Gudelia Figueroa Preciado y al Dr. Oscar Vega Amaya. Sus observaciones y sugerencias ayudaron a mejorar significativamente este trabajo.

De manera muy especial quiero agradecer a mis padres y mis hermanos por todo su apoyo tanto moral como económico, a mi esposa Beatriz Cecilia por motivarme a no parar, aún cuando se presentaron dificultades, y a mis hijas Cecilia Guadalupe y Beatriz Ixchel por su alegría y comprensión.

A la Universidad de Sonora, a los maestros del Departamento de Matemática por transmitir de manera muy profesional su conocimiento a los alumnos y a mis compañeros de la licenciatura por su amistad y solidaridad.

Por el apoyo económico que recibí por parte del Laboratorio de Investigación y Consultoría Estadística (LICE) del Departamento de Matemáticas de la Universidad de Sonora, a través de una beca patrocinada por la empresa Datos y Cifras del Noroeste S.C. durante la elaboración de este trabajo. También a la Comisión Nacional del Agua por facilitar la información requerida para la elaboración de esta tesis.

A todos ustedes muchas gracias.

Contenido

Prefacio	4
1 Introducción	7
1.1 Precipitación atmosférica	7
1.2 Inundaciones	8
1.2.1 Impacto social	9
1.2.2 Impacto económico	10
1.3 Precipitación en Hermosillo, Sonora	14
1.4 Gestión de riesgos	16
1.5 Caso de estudio: Precipitación máxima anual en Hermosillo, Sonora . .	17
2 Teoría de valores extremos	19
2.1 Introducción	19
2.2 Formulación del modelo	20
2.3 Distribuciones limite para valores extremos	22
2.4 Distribución de valores extremos generalizada	26
3 Conceptos de inferencia estadística	29
3.1 Teoría de estimación paramétrica vía el enfoque de verosimilitud	30
3.1.1 Función de verosimilitud	30
3.1.2 Función de verosimilitud relativa	34

3.1.3	Intervalos de verosimilitud	35
3.1.4	Intervalos de verosimilitud-confianza	37
3.1.5	Estimación en presencia de parámetros de estorbo: Verosimilitud perfil	39
3.2	Métodos de valuación de modelos	42
3.2.1	Métodos heurísticos	42
3.2.2	Uso de la verosimilitud perfil	45
3.2.3	Prueba de hipótesis	46
3.2.4	Pruebas de significancia: <i>p-valor</i>	48
4	Caso de estudio: Precipitación máxima anual en Hermosillo, Sonora	50
4.1	Descripción del problema	51
4.2	Parámetros de interés	52
4.2.1	Nivel de retorno	53
4.2.2	Vulnerabilidad a inundaciones	54
4.3	Modelación	54
4.4	Valoración de modelos: Weibull, Gumbel y Fréchet	55
4.5	Inferencias sobre parámetros de interés usando la verosimilitud perfil .	62
4.5.1	Nivel de retorno	64
4.5.2	Vulnerabilidad a inundaciones	66
4.6	Interpretación de resultados	68
4.6.1	Nivel de retorno	68
4.6.2	Vulnerabilidad a inundaciones	69
5	Conclusiones	72
	Bibliografía	73

Prefacio

La teoría de valores extremos es una rama de la estadística que se ha desarrollado bastante en las últimas dos décadas. Se encarga principalmente de analizar eventos asociados a los valores más altos o más bajos de una variable aleatoria. La teoría tiene aplicaciones en muchas áreas, principalmente en las asociadas al medio ambiente. Por ejemplo, un fenómeno ambiental es la precipitación atmosférica que se presenta principalmente en forma de lluvia, nieve, granizo o llovizna. De éstas, una de las más importantes y usuales es la lluvia la cual si se presenta de forma copiosa puede provocar inundaciones que impactan con gran fuerza a la sociedad. Por esta razón, es de interés saber el comportamiento de los niveles más altos de precipitación atmosférica, que se presentan en forma de lluvia; especialmente, para la gestión de riesgos en inundaciones, pues ésta se enfoca generalmente a reducir las pérdidas de vidas y la destrucción de propiedades e infraestructuras por causa de inundaciones.

El objetivo de este trabajo de tesis es modelar el nivel de precipitación máximo anual en Hermosillo, Sonora, y hacer inferencia estadística sobre parámetros de interés asociados al modelo. En particular, se hace inferencia sobre un parámetro ampliamente utilizado en teoría de extremos y otro propuesto en esta tesis llamados nivel de retorno y parámetro de vulnerabilidad a inundaciones, respectivamente. Pero más importante que todo esto, lo que se persigue es mostrar un proceso completo de inferencia estadística en un problema relevante de la sociedad.

A continuación se describen los aspectos más importantes abordados en cada capítulo.

En el Capítulo 1 se presentan conceptos preliminares como el de precipitación atmosférica e inundación. Se exhiben algunas situaciones que provocan las inundaciones básicamente en la salud. Se presentan algunas cifras que se dieron en el año 2008 correspondientes a pérdidas económicas por fenómenos hidrometeorológicos en México. Luego se documenta cronológicamente el impacto económico, social, etcétera de lo ocurrido después de algunas de las precipitaciones más altas que han ocurrido en Hermosillo, Sonora. También se determina en general en que consiste la gestión de riesgos y cuales son algunas de sus actividades. Por último se describe de forma breve el caso de estudio que se abordará en esta tesis.

En el Capítulo 2 se revisan algunos de los resultados más importantes de la teoría de valores extremos que serán utilizados a lo largo de esta tesis. Por su parte, el Capítulo 3 de esta tesis se encuentra dividido en dos partes. En la primera parte se define el método de máxima verosimilitud utilizado aquí para estimar los parámetros. Además, se definen conceptos importantes asociados a este enfoque: intervalos de verosimilitud-confianza, verosimilitud perfil, entre otros. La segunda parte corresponde a describir herramientas teóricas y heurísticas que son comúnmente empleadas para explorar y discernir modelos estadísticos apropiados, validarlos en el sentido de confrontación de los datos con un modelo dado (selección de modelos y bondad de ajuste), ayudar a seleccionar estadísticamente un modelo, dentro de un conjunto de modelos propuestos, para un fenómeno aleatorio de interés para el cual se cuenta con observaciones de un fenómeno aleatorio.

En el Capítulo 4 se aplican los conceptos definidos en los Capítulos 2 y 3 para modelar el nivel de precipitación máximo anual (NPMA) en Hermosillo, Sonora, valorar el modelo y hacer inferencia estadística sobre el parámetro llamado nivel de retorno y sobre otro parámetro propuesto en este trabajo llamado parámetro de vulnerabilidad a inundaciones. Aquí se ejemplifica de manera formal el proceso de inferencia estadística propuesto en este trabajo. Por ultimo, en el Capítulo 5 se presentan las conclusiones

generales de este trabajo.

Las principales aportaciones de este trabajo de tesis se comentan a continuación:

- Mostrar un proceso completo de inferencia estadística en un problema relevante de la sociedad: modelación, valoración del modelo, inferencia sobre parámetros de interés en el contexto del problema e interpretación de los resultados en el contexto del problema.
- Definición de conceptos relevantes en: Teoría de valores extremos, el enfoque de verosimilitud y valuación de modelos.
- Modelación del NPMA en Hermosillo, Sonora, e inferencia estadística sobre el nivel de retorno y otro parámetro, propuesto en este trabajo, llamado parámetro de vulnerabilidad a inundaciones.
- Revisión histórica del impacto social, económico y de salud de las inundaciones en Hermosillo, Sonora.

Posibles extensiones inmediatas de este trabajo:

- Explorar el uso de la distribución Gumbel para modelar el NPMA en otras regiones del estado de Sonora.
- Explorar el uso de modelos de teoría de valores extremos con parámetros que dependan del tiempo para modelar el NPMA.

Las simulaciones, cálculos de funciones de verosimilitud, optimizaciones y algunas figuras mostradas en esta tesis se realizaron con la ayuda del lenguaje de programación R versión 2.9.2. instalado en una computadora personal Pentium IV.

Capítulo 1

Introducción

En este capítulo se revisan conceptos preliminares como el de precipitación atmosférica y el de inundación, se exhiben algunas de las situaciones que durante y después de una inundación impactan considerablemente a la sociedad, se documentan algunas de las precipitaciones más copiosas que han ocurrido en la ciudad y que han originado inundaciones, se presenta a la gestión de riesgo como un enfoque estructurado para manejar la incertidumbre relativa a una amenaza y se menciona a la Comisión Nacional del Agua (CONAGUA) como la principal institución mexicana encargada en aplicarla en situaciones que involucren al agua. Por último, se describe brevemente el caso de estudio que se abordará en esta tesis.

1.1 Precipitación atmosférica

Se denomina *precipitación atmosférica* a todo tipo de partículas líquidas o sólidas de agua que caen desde las nubes y alcanzan la superficie de la tierra. Las precipitaciones atmosféricas más usuales son la lluvia, la nieve, el granizo, la llovizna y, en un sentido más restringido, el rocío y la escarcha, además de algunas variedades intermedias como el agua nieve o nieve líquida.

La precipitación atmosférica que se analiza en este trabajo es la lluvia, cuyo instrumento de medición es el *pluviómetro* el cual se encuentra fijo en un punto geográfico estratégico y mide la cantidad de agua precipitada en dicho lugar. Un esquema simple del funcionamiento del pluviómetro se muestra en la Figura 1.1.

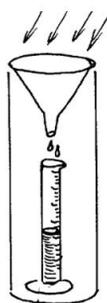


Figura 1.1: Ilustración de medición de la precipitación atmosférica usando un pluviómetro.

Por convención, la unidad de medida de una precipitación es el milímetro (mm) que equivale a un litro de agua distribuido uniformemente en una área de un metro cuadrado (*litro/m²*). Es decir, 1 *litro/m²* origina en una superficie completamente plana una lámina de agua de 1 mm de grosor.

La lluvia trae grandes beneficios a la flora y fauna de muchos lugares. El ser humano y muchas de sus actividades dependen directa e indirectamente del agua, la cual se obtiene en su mayoría de las lluvias. Sin embargo, las lluvias copiosas o atípicas combinadas con diferentes factores como la geografía del lugar, por ejemplo, pueden provocar inundaciones. En la siguiente sección se expondrá con más detalle el tema de inundación y su impacto en la sociedad.

1.2 Inundaciones

Una *inundación* se define de manera muy breve como la sumersión de áreas que no lo están normalmente. Cuando las inundaciones alcanzan lugares utilizados por el hom-

bre para habitar o para realizar sus actividades económicas, puede llegar a impactarlos considerablemente en lo moral, en lo económico y en la salud. La Organización Meteorológica Mundial (OMM) publicó en 2006 información acerca de la incidencia mundial de los desastres naturales y se observa que las inundaciones son las más frecuentes, véase la Figura 1.2.

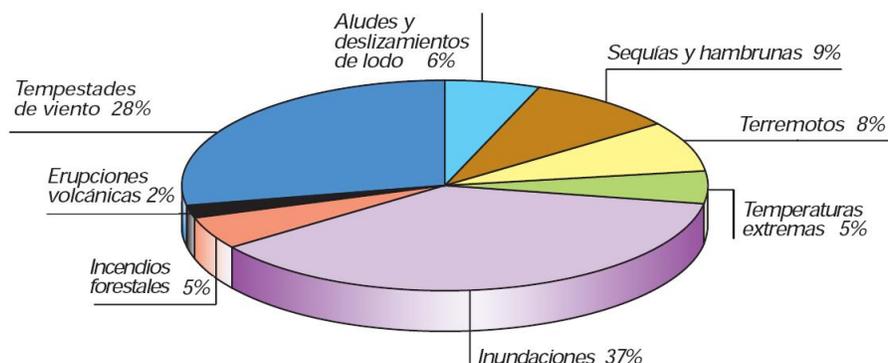


Figura 1.2: Incidencia mundial de los desastres naturales en 2006 según datos de la Organización Meteorológica Mundial (OMM).

1.2.1 Impacto social

Las inundaciones como todos los desastres naturales, traen situaciones que impactan profundamente a la sociedad. El deceso de un familiar o conocido, cambia radicalmente el entorno de quien sufre ésta pérdida, la frustración de perderlo todo en el hogar , la suspensión de los servicio básicos y en general situaciones que tienen que ver con la salud provocan gran angustia en la población.

Después de una inundación, las lesiones probablemente ocurren cuando las personas regresan a sus viviendas para limpiarlas. Se han presentado electrocuciones por líneas de alto voltaje caídas, manejo inapropiado de cables eléctricos y herramientas húmedas. Las lesiones por fuego y explosiones de gas también ocurren cuando se usan fósforos para inspeccionar los daños en estructuras oscuras. Las laceraciones y los pinchazos son

accidentes menores comunes en la remoción y limpieza de artículos con clavos y vidrio. También, el hábitat natural de muchos animales salvajes se puede ver alterado por las inundaciones. Como resultado, algunos animales, como las serpientes, se pueden ver forzadas a buscar refugio en las áreas que pueden estar habitadas o que son usadas por los humanos. De aquí que las mordeduras y ataques de animales son posibles.

En una inundación, existe el riesgo de exponerse a agentes químicos o biológicos peligrosos. Las tuberías subterráneas se pueden romper, los tanques de almacenamiento se pueden vaciar, los desechos tóxicos pueden fluir y los químicos almacenados en terrenos se pueden escapar. Los riesgos aumentan cuando las áreas industriales o agrícolas quedan sumergidas.

Personas que mueren a causa de la inundación y quedan expuestas (también el desalojo de las sepulturas en los cementerios inundados) afectan mentalmente a sus familiares. La contaminación de los pozos de agua después de extensas inundaciones ha llevado a la preocupación con respecto a la seguridad del agua de consumo humano.

Existe una gran cantidad de literatura psicosocial dedicada a la salud mental de las víctimas de las inundaciones. Principalmente se ha puesto atención en las consecuencias psicológicas de estos desastres sobre los individuos y las comunidades. Los estudios muestran que las alteraciones severas son raras, aunque los problemas emocionales leves y transitorios son comunes, véase Eric y Noji (1997).

1.2.2 Impacto económico

Además de afectar directamente a la población, las inundaciones afectan a los hospitales, las escuelas, los vehículos oficiales, la infraestructura eléctrica y petrolera, caminos y puentes, agua y alcantarillado. Todo esto repercute en la economía del país e indirectamente afectan la economía de los ciudadanos.

En México, la secretaría de gobernación (SEGOB), el sistema nacional de protección civil (SINAPROC) y el centro nacional de prevención de desastres (CENAPRED)

publican desde 2001 la serie de libros titulada “Impacto socioeconómico de los desastres en México” donde se puede consultar entre otras cosas los costos económicos de los desastres naturales ocurridos entre los años 1980 y 2008; en particular, los asociados con fenómenos hidrometeorológicos.

En esta serie para facilitar su análisis se agrupan a tales fenómenos en cuatro grandes rubros:

- Ciclones tropicales, lluvias e inundaciones
- Bajas temperaturas
- Sequías
- Otros (granizadas, heladas, fuertes vientos, tormentas eléctricas, etc.).

Por ejemplo, la población afectada en 2008 por este tipo de fenómenos fue de más de 1.5 millones de personas, se registraron 148 fallecimientos, asimismo, se presentaron daños en 60,400 viviendas, más de 650 escuelas, 146 instalaciones de salud (hospitales, centros de salud, clínicas, etc.), más de medio millón de hectáreas de diferentes cultivos y 1,811 kilómetros de carreteras (entre caminos rurales, alimentadores, carreteras federales y estatales), véase la tabla mostrada en la Figura 1.3.

Tipo de fenómeno	Muertos	Población afectada (personas) ^{1/}	Viviendas dañadas	Escuelas	Hospitales	Área de cultivo dañada y/o pastizales (ha)	Caminos afectados (km)	Total de daños (millones de pesos)	Porcentaje
Lluvias, inundaciones y ciclones tropicales	59	778,277	57,625	661	146	455,791.60	1,811.00	13,580.93	97.8
Bajas temperaturas	75	350,295	0	0	0	0.00	0.00	67.47	0.5
Sequías	0	40,021	0	0	0	102,194.60	0.00	142.53	1.0
Otros	14	340,484	2,775	2	0	1,888.90	0.00	99.22	0.7
Total	148	1,509,077	60,400	663	146	559,875.10	1,811.00	13,890.15	100.0

^{1/} Se refiere a población evacuada, herida o desaparecida.
Fuente: CENAPRED.

Figura 1.3: Resumen de daños ocasionados por fenómenos de origen hidrometeorológico durante el 2008.

Así, al igual que en 2007, los fenómenos que acumularon la mayor cantidad de daños fueron las lluvias, inundaciones y ciclones tropicales con 97.8% del total cuantificado.

El 2008 sobresale por el hecho de que, pese a no ser un año en el que hayan ocurrido desastres de grandes dimensiones, como las inundaciones en Hidalgo, Puebla, Veracruz y Tabasco en 1999, Isidore en 2002, los huracanes Emily, Stan y Wilma en el 2005 ó Dean y las inundaciones de Tabasco en 2007, sí se registraron cerca de 30 desastres con impactos focalizados que superaron los cien millones de pesos cada uno. Por tal motivo el monto de daños se acercó a los 14 mil millones de pesos, convirtiéndolo en el tercer año que mayores pérdidas económicas haya sufrido a consecuencia de desastres de origen hidrometeorológico en la última década (véase la Figura 1.4).

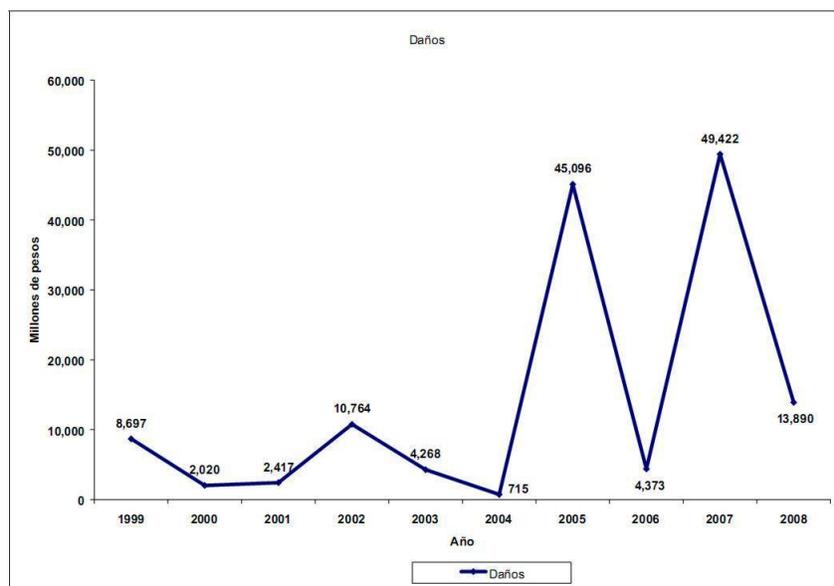


Figura 1.4: Distribución de daños ocasionados por fenómenos de origen hidrometeorológico en la última década. Fuente: CENAPRED.

En la tabla mostrada en la Figura 1.5 se presentan los daños ocasionados por lluvias, inundaciones y ciclones tropicales por estados ocurridos durante el 2008. Nótese que los estados de Tabasco, Veracruz, Oaxaca y Sonora concentraron cerca del 73.9% de las pérdidas económicas. Igualmente, que Sonora se encuentra en la cuarta posición

con un total de daños económicos de 902.5 millones de pesos y más de 26,000 personas afectadas.

Estado	Muertos	Población afectada (personas) ^{1/}	Viviendas dañadas	Escuelas	Hospitales	Área de cultivo dañada y/o pastizales (ha)	Caminos afectados (km)	Total de daños (millones de pesos)
Tabasco	1	41,800	2,525	456	8	360,022.0	965.5	4,596.0
Veracruz	8	397,743	26,441	64	98	43,210.5	385.2	3,502.2
Oaxaca	4	19,948	3,989	4	0	0.0	0.0	1,045.9
Sonora	4	26,113	3,024	53	36	3,370.0	460.3	902.5
Chihuahua	8	23,387	3,999	14	0	7,897.3	0.0	863.8
Tamaulipas	4	168,793	1,662	10	0	0.0	0.0	647.3
Baja California Sur	1	3,600	772	10	0	0.0	0.0	438.8
Durango	0	14,527	2,440	0	0	6,766.1	0.0	418.2
Nuevo León	1	858	171	0	0	0.0	0.0	210.9
Aguascalientes	2	6,194	180	9	0	0.0	0.0	151.1
San Luis Potosí	2	5,855	22	0	0	8,509.6	0.0	144.5
Sinaloa	1	11,621	1,778	0	0	7,169.1	0.0	142.8
Quintana Roo	0	975	195	0	0	0.0	0.0	141.6
Nayarit	0	3,235	0	0	0	4,071.4	0.0	108.7
Chiapas	0	33,747	4,238	29	0	3,845.1	0.0	89.8
Zacatecas	5	4,032	3,304	1	2	190.0	0.0	75.1
Tlaxcala	0	4,729	341	0	0	8,167.8	0.0	58.8
Michoacán	0	730	43	0	1	1,560.8	0.0	17.0
Jalisco	7	1,602	253	0	0	1,011.9	0.0	15.0
Guanajuato	3	1,499	866	4	1	0.0	0.0	3.9
Puebla	0	2,070	414	4	0	0.0	0.0	2.5
Guerrero	1	2,272	437	0	0	0.0	0.0	1.7
Estado de México	2	1,152	230	0	0	0.0	0.0	1.4
Morelos	0	305	61	0	0	0.0	0.0	0.5
Querétaro	0	875	175	3	0	0.0	0.0	0.5
Coahuila	0	320	20	0	0	0.0	0.0	0.2
Campeche	0	144	17	0	0	0.0	0.0	0.2
Baja California	1	11	2	0	0	0.0	0.0	0.1
Yucatán	0	100	20	0	0	0.0	0.0	0.1
Hidalgo	3	33	6	0	0	0.0	0.0	0.1
Colima	1	1	0	0	0	0.0	0.0	0.0
Distrito Federal	0	6	0	0	0	0.0	0.0	0.0
Total	59	778,277	57,625	661	146	455,791.6	1,811.0	13,580.9

^{1/} Se refiere a población evacuada, herida o desaparecida.

Fuente: CENAPRED.

Figura 1.5: Resumen de daños ocasionados por lluvias, inundaciones y ciclones tropicales durante el 2008 a nivel estatal.

1.3 Precipitación en Hermosillo, Sonora

Los sucesos que se describen a continuación fueron tomados de publicaciones posteriores a cada evento del periódico El Imparcial de Hermosillo, Sonora.

El 1 de septiembre de 1989 se presentó la lluvia más copiosa de ese año (100 mm). En aquella ocasión se suspendieron los servicios de energía eléctrica y agua potable en el norte de la ciudad. En el sur se presentaron inundaciones en las colonias Emiliano Zapata, Fraccionamiento Fovissste, Nacameri y Las Villas. En el centro de la ciudad el agua inundó comercios y a la central de autobuses de Hermosillo. Al norte de la ciudad varios cruces de terracería quedaron intransitables entre los que se destacan Guadalupe Victoria y López Portillo, Reforma y López Portillo, Monteverde y López Portillo y Arizona y López Portillo.

El 12 de agosto de 1990 llueven 112.5 mm y causan inundaciones en colonias del poniente, suroeste y noroeste de Hermosillo, así como en el ejido el Tronconal y varios campos de la Costa de Hermosillo. Se estima que miles de Hermosillenses fueron afectados. Decenas de automovilistas se quedaron varados por averías que causaron las corrientes de aguas en las calles. En el ejido el Tronconal, ubicado por la carretera a Ures, unas diez casas fueron totalmente cubiertas por el agua. En Hermosillo las colonias que fueron afectadas por inundaciones fueron: Los Arcos, Villa Satélite, Raquet Club, Las Granjas, Valle Verde, Palmar del Sol, Pimentel y Lomas de Madrid, así como cercanías a estas. También hubo calles que se inundaron por el alto flujo de agua como Periférico Poniente, Bulevar García Morales, Reforma, Norberto Aguirre, Bulevar Navarrete, Olivares, entre otras. Además, cientos de pasajeros que viajaban rumbo al sur se quedaron en la central de autobuses y en la estación del ferrocarril pues ni los camiones ni el tren pudieron continuar más allá de Hermosillo por el rompimiento de la vía y la carretera Hermosillo-Guaymas a causa de las fuertes lluvias. Como anécdota se recuerda la colocación de “agujas” en la cresta del vertedor de la presa Abelardo L. Rodríguez y su vigilancia las 24 horas para monitorear los niveles del agua y la

situación de la presa.

El 23 de agosto de 1992 se presentó en Sonora la tormenta tropical Lester provocando una precipitación de 142.0 mm en Hermosillo que causó inundaciones en las colonias El Apache, Las Villas y Las Minitas, donde se evacuaron a más de 100 familias; también en la colonia El Ranchito, Nacameri y algunas colonias del norte se presentaron inundaciones. Con esta precipitación se desbordó el río San Miguel e inundó gran parte del ejido La Victoria. Además se rompió el colector del Ranchito, provocando que las aguas negras se mezclaran con las del cauce natural del río Sonora.

El 11 de Noviembre de 1994 se presentaron en Hermosillo algunas inundaciones y descomposturas de automóviles por la precipitación de 152.6 mm. Se inundó parte del poblado El Tronconal, así como la mayor parte de las calles de la costa de Hermosillo: el Poblado Miguel Alemán, Kino Viejo y Kino Nuevo. Como anécdota se recuerda la caída de la torre y parte de la estructura de la iglesia de la Candelaria ubicada en la colonia Villa de Seris considerada monumento histórico de Sonora y cuya construcción data de 1850.

El 9 de octubre del 2000 se registró una precipitación de 125.5 mm que originó que alrededor de 15 colonias se vieran afectadas por inundación de casas. También algunos automovilistas sufrieron problemas al quedar atrapados en el interior de canales y arroyos. En las siguientes colonias hubo reporte de inundaciones: Mártires de Cananea, Carmen Serdán, Álvaro Obregón, Ladrilleras, Jesús García, Villas del Rey, Olivares, Balderrama, Sahuaro Final y San Jerónimo. Como anécdota no grata se recuerda que el canal Lázaro Cárdenas se desbordó y mucha del agua que traía se desfogó hacia el sur por calles como la Monteverde, Arizona y Reforma hasta el bulevar Luis Encinas que funcionó como recolector de los flujos del norte.

El 29 de julio del 2010 una precipitación pluvial de 80 mm en forma generalizada (en promedio) y la más alta en el aeropuerto Ignacio Pesqueira con 103 mm provocaron que alrededor de 46 colonias resultaran afectadas por inundaciones. Niveles de hasta

metro y medio se registraron en el fraccionamiento Tulipanes 1 y otras colonias como Las Villitas y San Juan por lo que hubo necesidad de evacuar a familias. Hubo además desbordamiento de canales y arroyos como el que se encuentra a un costado de la Universidad Kino, el arroyo Las Víboras ubicado al sur de la ciudad, entre otros.

A continuación se describirá un enfoque usado para buscar alternativas para disminuir, sobrellevar y responder de forma oportuna a tales situaciones llamado gestión de riesgo.

1.4 Gestión de riesgos

En general, la gestión de riesgos consiste en una serie de actividades diseñadas para reducir las pérdidas de vidas humanas y la destrucción de propiedades e infraestructuras. Los resultados de este proceso continuo de manejo o gestión de riesgos pueden ser divididos en:

- Medidas para disminuir el riesgo de desastres a largo plazo (**prevención**), eliminando sus causas como la intensidad de los fenómenos, la exposición o el grado de vulnerabilidad.
- Medidas de **preparación** cuyo objeto es asegurar una respuesta apropiada en caso de necesidad, incluyendo alertas tempranas oportunas y eficaces, así como evacuación temporal de gente y bienes de zonas amenazadas.
- Medidas de **respuesta** cuando está sucediendo o ha sucedido un desastre (manejo o gestión de desastres, recuperación, reconstrucción).

Véase COSUDE (2002).

Las medidas de prevención incluyen la realización de estudios y análisis para identificar, evaluar y cuantificar el nivel de amenaza, vulnerabilidad y riesgo, así como las acciones para mitigar (reducir) los efectos de los peligros observados. Los estudios y

análisis de identificación y evaluación de amenazas y vulnerabilidades están englobados en el denominado análisis de riesgos, el cual tiene como objetivo servir como base para la elaboración de los planes de reducción de desastres.

En México la Comisión Nacional del Agua (CONAGUA) es la organización que, entre otras de sus actividades, se encarga de aplicar la gestión de riesgos en situaciones que tengan que ver con el agua, en particular con las inundaciones.

Un enfoque para intentar conocer los riesgos reales de inundación en una determinada área geográfica es el análisis de los registros de niveles de precipitación máximos anuales. En este trabajo de tesis se modela el NPMA de Hermosillo, Sonora, usando un modelo de valores extremos generalizados. Además, se propone un índice de vulnerabilidad a inundaciones que será estimado junto con el nivel de retorno, a partir de datos reales registrados en esta localidad. Cabe señalar aquí que todo esto puede ayudar a tomar decisiones que permitan reducir la vulnerabilidad de la ciudad respecto a inundaciones.

1.5 Caso de estudio: Precipitación máxima anual en Hermosillo, Sonora

En la Sección 1.3 se mostró que a través del paso de los años la ciudad de Hermosillo, Sonora ha estado expuesta a inundaciones. En esta tesis se analizarán datos reales de precipitación máxima anual registradas en estaciones meteorológicas de la CONAGUA ubicadas en la ciudad de Hermosillo Sonora desde 1965 hasta 2009. Los datos utilizados en este trabajo fueron proporcionados por el Ing. Martín Barrón Félix jefe del departamento meteorología de la CONAGUA en Sonora.

El objetivo principal de analizar dichos datos es ejemplificar un proceso completo de inferencia estadística (modelación, valoración del modelo, inferencia sobre parámetros de interés en el contexto del problema e interpretación de los resultados en el contexto

del problema) y presentar las estimaciones del nivel de retorno y el parámetro de vulnerabilidad a inundaciones que tiene Hermosillo. Los resultados de este caso de estudio se muestran de manera detallada en el Capítulo 4.

A continuación, en los Capítulos 2 y 3, se presentan conceptos teóricos relevantes de la teoría de extremos y de inferencia estadística que serán utilizados en esta tesis.

Capítulo 2

Teoría de valores extremos

2.1 Introducción

La teoría de valores extremos a emergido como una de las más importantes ramas de la estadística que se aplica a la ciencia en los últimos 50 años; en particular en las últimas dos décadas se ha desarrollado rápidamente tanto desde el punto de vista metodológico como de las aplicaciones.

La mayoría de los estudios estadísticos tratan de la modelación del promedio de la distribución de la variable de interés, el cual se estima con la media muestral. En la teoría de valores extremos el interés principal no está en el promedio, sino en los valores más bajos o más altos de la variable bajo estudio, es decir, el interés está en los eventos asociados a la cola de la distribución.

Por ejemplo en estudios de oceanografía es necesario estudiar el comportamiento de corrientes marinas extremas, en estadística ambiental es necesario analizar niveles altos de ozono en determinada región, en climatología es necesario conocer el comportamiento de velocidades extremas de huracanes, en ingeniería el aumento del flujo de un río, en telecomunicaciones es necesario predecir el tráfico en las líneas de comunicación, en finanzas un gran decremento o aumento del valor de una acción, en el

mercado valores máximos o mínimos en un portafolio de inversión, etc.

Cualquier tipo de pregunta relacionada con las situaciones antes mencionadas tendrá que involucrar un manejo adecuado de la información con que se cuente y de una aplicación cuidadosa de herramientas analíticas para llevar a cabo el proceso de análisis e inferencia.

2.2 Formulación del modelo

En esta sección se desarrolla el modelo que representa la piedra angular de la teoría de valores extremos. El modelo se centra en el comportamiento estadístico de

$$M_n = \max(X_1, \dots, X_n),$$

donde X_1, \dots, X_n es una sucesión de variables aleatorias independientes e idénticamente distribuidas (vaid) con función de distribución F . En las aplicaciones, las X_i usualmente representan valores de un proceso de medición en una escala de tiempo regular. Por ejemplo, mediciones por hora del nivel del mar, mediciones por día de la temperatura media, mediciones por día del nivel de precipitación, etcétera. De manera que M_n representa el máximo sobre n unidades de tiempo dentro del proceso de observación. En particular si n es el número de observaciones en el año, entonces M_n corresponde al máximo anual.

En teoría, se puede obtener la distribución de M_n en forma exacta para todos los valores de n ,

$$\begin{aligned} P(M_n \leq z) &= P(X_1 \leq z, \dots, X_n \leq z) \\ &= P(X_1 \leq z) \times \dots \times P(X_n \leq z) \\ &= [F(z)]^n. \end{aligned} \tag{2.1}$$

Sin embargo, esto no es muy útil en la práctica ya que usualmente F es desconocida. Se podría estimar F usando los datos observados y entonces sustituir esta estimación

en (2.1). Desafortunadamente pequeñas discrepancias en la estimación de F pueden producir grandes diferencias para F^n .

Un enfoque alternativo es aceptar que F es desconocida y buscar aproximar familias de modelos a la distribución de F^n que se pueda estimar sobre la base de los datos extremos solamente. Esto es similar a la práctica habitual de la aproximación de una muestra de promedios a través de la distribución normal, justificada por el teorema de límite central. Los argumentos en este capítulo son esencialmente análogos al de teorema de límite central solamente que para una muestra de valores extremos.

Sin embargo, para lograr esto no es suficiente considerar el comportamiento de F^n cuando n tiende a infinito ($n \rightarrow \infty$), ya que si se define u_F como el valor más pequeño para z tal que $F(z) = 1$, entonces

$$\lim_{n \rightarrow \infty} F^n(z) = \delta(z) = \begin{cases} 0 & \text{si } z < u_F, \\ 1 & \text{si } z \geq u_F. \end{cases}$$

Es decir, F^n tiende en distribución, cuando $n \rightarrow \infty$, a la distribución de la variable aleatoria degenerada $\delta(z)$ que toma el valor u_F con probabilidad 1. Utilizar esta distribución concentrada en un punto para aproximar la distribución de M_n no resulta informativo. Esta dificultad se evita al permitir una normalización lineal de la variable M_n ,

$$M_n^* = \frac{M_n - b_n}{a_n},$$

para sucesiones de constantes $\{a_n > 0\}$ y $\{b_n\}$.

En lo que resta del capítulo se presentarán algunos de los resultados más importantes de la teoría de valores extremos que surgieron al considerar esta normalización de M_n .

2.3 Distribuciones limite para valores extremos

El siguiente teorema tomado de Embrechts et al. (1997, p. 121), fue propuesto en 1928 por Fisher y Tippett y probado rigurosamente por Gnedenko en 1943. Este teorema permite aproximar la distribución de una normalización lineal de M_n .

Teorema 2.3.1 (Fisher y Tippett) *Suponga que existen sucesiones de constantes $\{a_n > 0\}$ y $\{b_n\}$ tales que*

$$P \left[\frac{(M_n - b_n)}{a_n} \leq z \right] = F^n(a_n z + b_n) \rightarrow G(z) \text{ cuando } n \rightarrow \infty, \quad (2.2)$$

donde G es una distribución que no está concentrada en un punto (no degenerada). Entonces G pertenece a una de las siguientes tres familias de distribuciones:

$$I: G(z) = \exp \left\{ - \exp \left[- \left(\frac{z - \mu}{\sigma} \right) \right] \right\}, \text{ si } z \in \mathbb{R};$$

$$II: G(z) = \begin{cases} 0, & \text{si } z \leq \mu, \\ \exp \left[- \left(\frac{z - \mu}{\sigma} \right)^{-\beta} \right], & \text{si } z > \mu; \end{cases}$$

$$III: G(z) = \begin{cases} \exp \left\{ - \left[- \left(\frac{z - \mu}{\sigma} \right)^\beta \right] \right\}, & \text{si } z < \mu, \\ 1, & \text{si } z \geq \mu, \end{cases}$$

donde $\sigma > 0$, $\mu \in \mathbb{R}$, y $\beta > 0$ son parámetros.

Las distribuciones asintóticas del Teorema 2.3.1 son llamadas *Distribuciones de Valores Extremos* (DVE) y de forma individual las familias *I*, *II* y *III* son conocidas como la familia Gumbel, la familia Fréchet y la familia Weibull, respectivamente.

El Teorema 2.3.1 implica que, cuando M_n pueda ser estabilizada con sucesiones $\{a_n\}$ y $\{b_n\}$ adecuadas, la correspondiente normalización de la variable M_n^* tiene una distribución limite que debe de ser una de los tres tipos de distribución para valores extremos. La parte esencial de este resultado es que los tres tipos de distribuciones *I*,

II y *III* son los únicos límites posibles para la distribución de M_n^* , independientemente de la distribución F de la población. A continuación se presentan ejemplos particulares donde es posible ver que se cumple tal afirmación.

Ejemplo 2.3.2 *Supóngase que X_1, X_2, \dots, X_n es una sucesión de v.a.i.i.d con función de distribución exponencial y parámetro $\lambda = 1$, $F(x) = 1 - \exp(-x)$, para $x > 0$. Ahora, considerése $a_n = 1$ y $b_n = \log n$. Entonces,*

$$\begin{aligned} P(M^* \leq z) &= P\left(\frac{M_n - b_n}{a_n} \leq z\right) \\ &= P(M_n \leq a_n z + b_n) \\ &= P(M_n \leq z + \log n). \end{aligned}$$

Luego, de (2.1) se sigue que

$$\begin{aligned} P(M^* \leq z) &= [F(z + \log n)]^n \\ &= \{1 - \exp[-(z + \log n)]\}^n \\ &= [1 - \exp(-z) \cdot \exp(-\log n)]^n \\ &= \{1 - \exp(-z) \cdot \exp[\log(n^{-1})]\}^n \\ &= [1 - n^{-1} \cdot \exp(-z)]^n. \end{aligned} \tag{2.3}$$

Entonces de (2.3) se tiene que

$$\lim_{n \rightarrow \infty} P(M^* \leq z) = \exp[-\exp(-z)],$$

para cada valor fijo de $z \in \mathbb{R}$. Por lo tanto, con las sucesiones a_n y b_n elegidas, la distribución límite para M_n cuando $n \rightarrow \infty$ es la distribución Gumbel la cual pertenece a la familia de DVE.

Ejemplo 2.3.3 *Supóngase que X_1, X_2, \dots, X_n es una sucesión de v.a.i.i.d con función de distribución Fréchet, $F(x) = \exp(-\frac{1}{x})$, para $x > 0$. Ahora, considerése $a_n = n$ y*

$b_n = 0$. Entonces,

$$\begin{aligned} P(M^* \leq z) &= P\left(\frac{M_n - b_n}{a_n} \leq z\right) \\ &= P(M_n \leq a_n z + b_n) \\ &= P(M_n \leq nz). \end{aligned}$$

Luego, de (2.1) se sigue que

$$\begin{aligned} P(M^* \leq z) &= [F(nz)]^n \\ &= \left[\exp\left(-\frac{1}{nz}\right)\right]^n \\ &= \left[\exp\left(-\frac{1}{z}\right)\right]^n, \end{aligned} \tag{2.4}$$

para todo valor de n y para cada valor fijo $z > 0$. Por lo tanto, en este caso la distribución límite es la misma distribución Fréchet que también pertenece a la familia de DVE.

Ejemplo 2.3.4 Supóngase que X_1, X_2, \dots, X_n es una sucesión de vaíid con función de distribución uniforme de parámetros $a = 0$ y $b = 1$. Para cada valor fijo $z < 0$, supóngase que $n > -z$ y considerése $a_n = 1/n$ y $b_n = 1$. Entonces,

$$\begin{aligned} P(M^* \leq z) &= P\left(\frac{M_n - b_n}{a_n} \leq z\right) \\ &= P(M_n \leq a_n z + b_n) \\ &= P(M_n \leq n^{-1}z + 1). \end{aligned}$$

Luego, de (2.1) se sigue que

$$\begin{aligned} P(M^* \leq z) &= [F(n^{-1}z + 1)]^n \\ &= \left(1 + \frac{z}{n}\right)^n. \end{aligned} \tag{2.5}$$

Entonces de (2.5) se tiene que

$$\lim_{n \rightarrow \infty} P(M^* \leq z) = \exp(z).$$

Por lo tanto, la distribución límite es una distribución del tipo Weibull que es una de las DVE.

En este sentido, el teorema de Fisher y Tippett se puede usar para justificar la suposición de que M_n tiene una DVE.

Las densidades asociadas a las DVE para máximos se obtienen de manera directa del Teorema 2.3.1 derivando respectivamente cada distribución y son las siguientes:

$$\text{Gumbel: } \lambda(z; \mu, \sigma) = \frac{1}{\sigma} \exp \left[-\exp \left(\frac{\mu - z}{\sigma} \right) + \frac{\mu - z}{\sigma} \right], \quad z \in \mathbb{R};$$

$$\text{Fréchet: } \phi(z; \mu, \sigma, \beta) = \frac{\beta}{\sigma} \left(\frac{z - \mu}{\sigma} \right)^{-(1+\beta)} \exp \left[-\left(\frac{z - \mu}{\sigma} \right)^{-\beta} \right], \quad z \geq \mu;$$

$$\text{Weibull: } \psi(z; \mu, \sigma, \beta) = \frac{\beta}{\sigma} \left(\frac{\mu - z}{\sigma} \right)^{\beta-1} \exp \left[-\left(\frac{\mu - z}{\sigma} \right)^{\beta} \right], \quad z \leq \mu.$$

Las expresiones para el cuantil de probabilidad acumulada p de estas distribuciones son:

$$\text{Gumbel: } Q_p = \mu - \sigma \{ \log [-\log (p)] \},$$

$$\text{Fréchet: } Q_p = \mu + \sigma \left\{ [-\log (p)]^{-\frac{1}{\beta}} \right\},$$

$$\text{Weibull: } Q_p = \mu - \sigma \left\{ [-\log (p)]^{\frac{1}{\beta}} \right\}.$$

Nótese que estos cuantiles se obtienen al resolver la ecuación $G(Q_p) = p$, para $p \in (0, 1)$, donde G es la función de distribución Gumble, Fréchet y Weibull, respectivamente.

En aplicaciones de la teoría de valores extremos es usual adoptar una de las tres familias, Gumbel, Fréchet ó Weibull, para modelar eventos extremos, y luego estimar los parámetros relevantes asociados a la distribución elegida. Desafortunadamente el Teorema 2.3.1 no proporciona información sobre cómo discernir entre las tres familias. A continuación se presenta una reformulación del Teorema 2.3.1 que permite principalmente la aplicación de técnicas de inferencia estadística para cuantificar la incertidumbre a la hora de elegir una de ellas.

2.4 Distribución de valores extremos generalizada

R. Von mises en 1954 y A. F. Jenkinson en 1955 propusieron una reformulación del Teorema 2.3.1 en la cual combinan las familias de distribuciones Gumbel, Fréchet y Weibull en una única familia de distribuciones

$$G(z) = \exp \left\{ - \left[1 + c \left(\frac{z - a}{b} \right) \right]^{-\frac{1}{c}} \right\}, \quad (2.6)$$

definida para z tales que $1 + c \left(\frac{z - a}{b} \right) > 0$, donde $a \in \mathbb{R}$, $b > 0$ y $c \in \mathbb{R}$.

La distribución (2.6) es conocida como *Distribución de Valores Extremos Generalizada* (DVEG). La distribución tiene tres parámetros, un parámetro de localización a , un parámetro de escala b y un parámetro de forma c . En esta reparametrización, la distribución Fréchet corresponde al caso $c > 0$, la distribución Weibull al caso $c < 0$ y la distribución Gumbel al caso $c = 0$, que se interpreta como el límite de (2.6) cuando $c \rightarrow 0$.

Las relaciones de los parámetros de la DVEG con los parámetros de las DVE son las siguientes:

Gumbel: $\sigma = b$, $\mu = a$;

Fréchet: $\beta = \frac{1}{c}$, $\sigma = \frac{b}{c}$, $\mu = a - \frac{b}{c}$;

Weibull: $\beta = \frac{-1}{c}$, $\sigma = \frac{-b}{c}$, $\mu = a - \frac{b}{c}$.

Así, el Teorema 2.3.1 puede ser reformulado de la siguiente forma.

Teorema 2.4.1 *Suponga que existen sucesiones de constantes $\{a_n > 0\}$ y $\{b_n\}$ tales que*

$$P \left[\frac{(M_n - b_n)}{a_n} \leq z \right] = F^n(a_n z + b_n) \rightarrow G(z) \text{ cuando } n \rightarrow \infty, \quad (2.7)$$

donde G es una distribución que no está concentrada en un punto. Entonces, G es un miembro de la familia DVEG

$$G(z) = \exp \left\{ - \left[1 + c \left(\frac{z-a}{b} \right) \right]^{-\frac{1}{c}} \right\}, \quad (2.8)$$

para z tales que $1 + c \left(\frac{z-a}{b} \right) > 0$, donde $a \in \mathbb{R}, b > 0$ y $c \in \mathbb{R}$.

La densidad del modelo DVEG con parámetros (a, b, c) y $c \neq 0$, está dada por

$$g(z; a, b, c) = \frac{1}{b} \left[1 + c \left(\frac{z-a}{b} \right) \right]^{-\left(1+\frac{1}{c}\right)} \exp \left\{ - \left[1 + c \left(\frac{z-a}{b} \right) \right]^{-\frac{1}{c}} \right\},$$

para z tales que $1 + c \left(\frac{z-a}{b} \right) > 0$.

Cuando $c = 0$ la densidad está dada por

$$g(z; a, b) = \frac{1}{b} \exp \left[- \exp \left(\frac{a-z}{b} \right) + \frac{a-z}{b} \right], \quad (2.9)$$

para $z \in \mathbb{R}$.

Nótese que el soporte de la función de densidad de la DVEG depende de los tres parámetros (a, b, c) , mientras que para las familias Weibull y Fréchet sólo depende de un parámetro μ al cual se le llama parámetro umbral.

El cuantil de probabilidad acumulada p de la DVEG está dado por la expresión

$$Q_p = \begin{cases} a - \frac{b}{c} \{1 - [-\log(p)]^{-c}\} & \text{si } c \neq 0, \\ a - b \{\log[-\log(p)]\}, & \text{si } c = 0. \end{cases}$$

Sin duda, contar con la DVEG es de gran utilidad pues con solo estimar uno de sus parámetros (el parámetro c), se obtiene información acerca de cual modelo utilizar. Además, permite aplicar técnicas de inferencia estadística para cuantificar la incertidumbre al elegir un modelo.

Es importante mencionar que así como son de interés las distribuciones para máximos también lo pueden ser las de mínimos. Los resultados para mínimos son inmediatos, basta considerar los correspondientes resultados para M_n . Si $Y_i = -X_i$, para

$i = 1, \dots, n$, $\tilde{M}_n = \min \{X_1, \dots, X_n\}$ y $M_n = \max \{Y_1, \dots, Y_n\}$, entonces $\tilde{M}_n = -M_n$.

Por lo tanto, para n grande,

$$\begin{aligned} P(\tilde{M}_n \leq z) &= P(-M_n \leq z) \\ &= P(M_n \geq -z) \\ &= 1 - P(M_n \leq -z). \end{aligned} \tag{2.10}$$

Así, a partir de (2.10) es posible deducir la DVEG para mínimos y luego reformular el Teorema 2.3.1 y 2.4.1.

En el siguiente capítulo se describe el método utilizado en esta tesis para la estimación de parámetros; así como varias técnicas estadísticas para la valuación de modelos.

Capítulo 3

Conceptos de inferencia estadística

El objetivo de un modelo de probabilidad es representar alguna faceta de la realidad a través de una abstracción. Un modelo permite manipular artificialmente aspectos de la realidad, con el objeto de obtener respuestas. En una aplicación de la estadística, existen pues, dos mundos: el de la realidad, y el del modelo de probabilidad que lo representa. En el mundo real típicamente existe formulada alguna pregunta de interés. Si la respuesta a la pregunta puede obtenerse manipulando la realidad, entonces no es útil ni necesario el concepto de un modelo de probabilidad. Con un modelo de probabilidad, podemos estudiar la realidad con una representación abstracta. En la medida en que el modelo de probabilidad represente adecuadamente a la realidad, las respuestas que obtengamos con el modelo serán también apegadas a la realidad.

Dado un modelo de probabilidad paramétrico para un fenómeno aleatorio de interés, cualquier proceso de inferencia estadística debe contemplar al menos los siguientes pasos: (véase Cox y Hinkley 1974)

- Validación del modelo elegido.
- Estimación de los parámetros involucrados en el modelo.

En esta tesis se propone usar el siguiente proceso de inferencia estadística:

1. Modelación.
2. Valoración del modelo.
3. Inferencia sobre parámetros de interés en el contexto del problema.
4. Interpretación de los resultados en el contexto del problema.

Para abordar los puntos 2 y 3 antes mencionados se describirá en la siguiente sección la teoría de estimación paramétrica vía el enfoque de verosimilitud y posteriormente se presentarán algunos métodos estadísticos tanto teóricos como heurísticos que proporcionan evidencia acerca de la correcta valuación de un modelo propuesto para los datos bajo estudio.

3.1 Teoría de estimación paramétrica vía el enfoque de verosimilitud

Un enfoque que aborda el problema de la estimación de parámetros, dentro de la teoría de estimación paramétrica y que es ampliamente utilizado por su objetivo uso de la información y por proporcionar resultados estadísticamente eficientes es el de verosimilitud. A continuación se describen conceptos fundamentales de esta teoría que se utilizarán en capítulos posteriores.

3.1.1 Función de verosimilitud

La *función de verosimilitud* o simplemente *verosimilitud* se define como la probabilidad de observar la muestra; pero como función de los parámetros involucrados en el modelo que haya sido seleccionado para el fenómeno aleatorio de interés. En particular, cuando se tiene una muestra de variables aleatorias discretas $X=(X_1, \dots, X_n)$ independientes

e idénticamente distribuidas con función de probabilidad $P(\cdot; \theta)$, la verosimilitud de θ se define como:

$$L(\theta; x) = \prod_{i=1}^n P(x_i; \theta), \quad (3.1)$$

donde $x=(x_1, \dots, x_n)$ representa a la muestra observada y θ es un vector conformado por un número finito de parámetros reales desconocidos $\theta=(\theta_1, \dots, \theta_m) \in \Theta \subset \mathbb{R}^m$, donde Θ es el *espacio parametral* o la región de valores posibles que puede tomar el vector de parámetros θ .

Fisher (1921) definió a la verosimilitud como proporcional a probabilidades, quizás con la intención de simplificar matemáticamente los cálculos asociados a esta metodología o quizás para enfatizar la importancia de la razón o cociente de verosimilitudes como una medida para discriminar entre dos valores diferentes del parámetro dada una muestra observada. Estos puntos serán retomados y explicados con más detalle en las secciones 3.1.2 y 3.1.3.

Por otro lado, cuando se tienen variables aleatorias continuas (iid) con función de densidad $f(x; \theta)$ y una precisión $h/2 > 0$ (pequeña) del instrumento de medición alrededor de las observaciones, la verosimilitud se define como en (3.1) escribiendo $P(x_i; \theta)$ como:

$$P(x_i; \theta) = P\left(x_i - \frac{h}{2} \leq X_i \leq x_i + \frac{h}{2}\right) = \int_{x_i - \frac{h}{2}}^{x_i + \frac{h}{2}} f(x; \theta) dx,$$

véase Barnard y Sprott (1983).

NOTA: Es importante mencionar aquí que muchos libros de texto definen a la verosimilitud en el caso continuo como proporcional al producto de densidades. Esto lo hacen puesto que una aproximación a la verosimilitud de θ dada en (3.1) es

$$L(\theta; x) \approx L_C(\theta; x) = \prod_{i=1}^n hf(x_i; \theta) \propto \prod_{i=1}^n f(x_i; \theta), \quad (3.2)$$

véase Montoya (2008).

En esta tesis específicamente se empleará la aproximación (3.2) puesto que es adecuada y como se verá más adelante en el Capítulo 4, no produce problemas inferenciales como los discutidos por Kalbfleisch (1985, Sección 9.4), Edwards (1992, Pág. 6, Pág. 167), Lindsey (1998), Sprott (2000, Pág. 19, Págs. 203-294), Lawless (2003, Pág. 186), Meeker y Escobar (1998, Pág. 275), Montoya *et al.* (2007, Págs. 195-202), entre otros.

Plausibilidad

La función de verosimilitud $L(\theta; x)$ permite ordenar la credibilidad o plausibilidad entre los valores de θ a la luz de los datos. Si $L(\theta_1; x) > L(\theta_2; x)$ entonces de (3.1) se sigue que la muestra observada x es más probable cuando el parámetro θ toma el valor θ_1 que cuando toma el valor θ_2 . Así, el cociente de verosimilitudes

$$\frac{L(\theta_1; x)}{L(\theta_2; x)},$$

es una medida de la plausibilidad de θ_1 relativa a θ_2 basada en la muestra observada x . El cociente $L(\theta_1; x)/L(\theta_2; x) = k$ significa que el valor θ_1 es k veces más plausible que el valor θ_2 en el sentido de que θ_1 hace a la muestra observada k veces más probable de lo que lo hace θ_2 .

Estimador de máxima verosimilitud

El *estimador de máxima verosimilitud* (EMV) de θ , es el valor del parámetro en el espacio parametral, $\hat{\theta} \in \Theta$, que satisface

$$L(\hat{\theta}; x) = \sup_{\theta \in \Theta} L(\theta; x). \quad (3.3)$$

Nótese que $0 \leq L(\theta; x) \leq 1$ debido a que es una probabilidad. Así, el supremo de $L(\theta; x)$ en $[0, 1]$ existe y es finito. Sin embargo, puede darse el caso en que el EMV no existe, o si existe, puede que no sea único. En lo que sigue se actuará como si existiera un único valor $\hat{\theta} \in \Theta$ que maximiza a la función de verosimilitud $L(\theta; x)$.

El EMV del parámetro θ es el valor más plausible de θ . Es decir, el EMV $\hat{\theta}$ es el valor de θ que explica mejor a la muestra observada en el sentido de que maximiza su probabilidad bajo el modelo de probabilidad propuesto para el fenómeno aleatorio de interés.

Cabe hacer notar que algunas veces, encontrar el valor de θ que maximiza $L(\theta; x)$ puede ser complicado debido a la forma que toma el producto de probabilidades y en consecuencia la verosimilitud. Usualmente es conveniente y válido trabajar con la *función de log-verosimilitud* de θ definida como el logaritmo natural de $L(\theta; x)$, es decir,

$$l(\theta; x) = \log [L(\theta; x)] = \sum_{i=1}^n \log [P(x_i; \theta)]. \quad (3.4)$$

Pues muchas veces es más fácil realizar procesos de optimización (maximización-minimización) cuando la función objetivo es una suma de funciones que cuando la función objetivo es un producto de funciones.

Es importante señalar que definir a la verosimilitud como proporcional a probabilidades [no igual como en (3.1)] no afecta el resultado de la optimización; sin embargo, es posible simplificar expresiones que resultan irrelevantes; quizá esta sea una de las razones por la cual Fisher (1921) la definió proporcional.

Invarianza funcional

En muchas ocasiones se utilizan reparametrizaciones de un modelo de probabilidad por conveniencia matemática, conveniencia computacional, propio interés de un parámetro que es función de otro que aparece en el modelo, etc. En estas situaciones, la invarianza funcional es una propiedad muy conveniente de la verosimilitud y significa que, en términos de plausibilidad, cualquier declaración cuantitativa acerca de θ implica una declaración cuantitativa correspondiente acerca de cualquier función uno a uno de θ . Una función $g(\theta)$ se llama función uno a uno si para cada valor de θ existe un valor único $g(\theta)$ y viceversa.

Teorema 3.1.1 *Supóngase que $L(\theta; x)$ es la función de verosimilitud de $\theta \in \Theta$. Sea $g(\theta) : \Theta \rightarrow \Delta$ una función uno a uno. Si*

$$\frac{L(\theta_1; x)}{L(\theta_2; x)} = k$$

entonces

$$\frac{L(\delta_1; x)}{L(\delta_2; x)} = k,$$

donde $\delta_1 = g(\theta_1)$ y $\delta_2 = g(\theta_2)$.

La demostración del teorema anterior se sigue por directa sustitución algebraica.

Una consecuencia inmediata de la propiedad de invarianza de la función de verosimilitud $L(\theta; x)$ es la invarianza del EMV de θ .

Teorema 3.1.2 (Invarianza del EMV) *Si $g(\theta) : \Theta \rightarrow \Delta$ es una función uno a uno y $\hat{\theta}$ es el EMV de θ entonces el EMV de $\delta = g(\theta)$ es $\hat{\delta} = g(\hat{\theta})$.*

Los detalles de la demostración se pueden ver en Casella y Berger (1990).

Un ejemplo simple de la propiedad de invarianza de la verosimilitud es el siguiente. Si $\theta > 0$ y $\delta = \log \theta$, entonces la verosimilitud del nuevo parámetro δ es $L(\delta; x) = L[\theta = \exp(\delta); x]$. Como consecuencia se tiene que el EMV de δ es $\hat{\delta} = \log \hat{\theta}$.

Esta propiedad permitirá que en el Capítulo 4 se puedan realizar inferencias sobre un parámetro llamado nivel de retorno y sobre otro parámetro propuesto en este trabajo denominado parámetro de vulnerabilidad a inundaciones, descritos más adelante, considerados de importancia para la gestión de riesgos.

3.1.2 Función de verosimilitud relativa

La función de verosimilitud relativa $R(\theta; x) : \Theta \rightarrow [0, 1]$ se define como

$$R(\theta; x) = \frac{L(\theta; x)}{\sup_{\theta \in \Theta} L(\theta; x)}, \quad (3.5)$$

donde $L(\theta; x)$ es la función de verosimilitud de θ .

Esta función está bien definida ya que el supremo de $L(\theta; x)$ en $[0, 1]$ existe y es finito. Además nótese que $R(\theta; x)$ yace entre cero y uno, para todo valor de θ en el espacio parametral. Valores de θ con $R(\theta; x)$ cercanos a uno son muy razonables o creíbles mientras que valores de θ con $R(\theta; x)$ cercanos a cero son poco creíbles a la luz de los datos.

La función de verosimilitud relativa se puede utilizar para hacer declaraciones cuantitativas acerca del grado sobre el cual valores del parámetro describen mejor el fenómeno aleatorio de interés, con base en la muestra observada.

En la práctica, generalmente es posible encontrar un único valor de θ en el espacio parametral que maximiza la función de verosimilitud $L(\theta; x)$. Cuando esto sucede entonces la función de verosimilitud relativa dada en (3.5) se puede escribir como la estandarización de la verosimilitud con respecto a la verosimilitud evaluada en dicho valor. Es decir, cuando el EMV del parámetro θ , $\hat{\theta}$, existe y es único entonces la función de verosimilitud relativa de θ es

$$R(\theta; x) = \frac{L(\theta; x)}{L(\hat{\theta}; x)}. \quad (3.6)$$

La función de verosimilitud relativa dada en (3.6) proporciona la plausibilidad de cualquier valor especificado de θ relativo al máximo verosímil $\hat{\theta}$, a la luz de los datos. Es decir, valores de θ con $R(\theta; x)$ cercanos a $R(\hat{\theta}; x) = 1$ hacen a la muestra observada casi tan probable como lo hace el EMV $\hat{\theta}$. En contraste, valores de θ con $R(\theta; x)$ cercanos a cero hacen que la probabilidad de la muestra observada sea pequeña con respecto a su máxima probabilidad alcanzada en $\hat{\theta}$.

3.1.3 Intervalos de verosimilitud

Una *región de verosimilitud de nivel c* para θ se define como

$$IV(c) = \{\theta : R(\theta; x) \geq c\}, \quad c \in [0, 1].$$

Cuando la región de verosimilitud es un intervalo entonces $IV(c)$ se llama intervalo de verosimilitud de nivel c para θ .

Los intervalos o regiones de verosimilitud indican los valores de θ que son más plausibles a un nivel c , a la luz de la muestra observada. Esto es, cualquier valor de θ en el $IV(c)$ tendrá verosimilitud relativa igual o mayor que c y cualquier valor de θ fuera del $IV(c)$ tendrá verosimilitud relativa menor que c . Esta es la forma en que los $IV(c)$ separan los valores plausibles de θ de los no plausibles a un nivel c (Sprott, 2000, pág. 14).

En particular cuando θ es de dimensión uno, una región de verosimilitud para un determinado nivel de plausibilidad c puede ser un intervalo o la unión de intervalos disjuntos, esto dependerá de la forma que tome la función de verosimilitud. El $IV(c)$ se puede obtener dibujando una línea horizontal en la gráfica de $R(\theta; x)$, paralela al eje cartesiano de θ y a una distancia c , y coleccionando todos los valores de θ cuya verosimilitud relativa $R(\theta; x)$ se encuentra por encima de dicha línea. Cuando se varía c desde 0 hasta 1 se obtiene una familia de intervalos de verosimilitud jerarquizados y anidados que convergen al EMV $\hat{\theta}$ cuando c tiende a 1. Nótese que el EMV $\hat{\theta}$ de θ , está contenido en todos los intervalos de verosimilitud porque $R(\hat{\theta}; x) = 1$. Entonces, esta familia de intervalos son equivalentes a la función de verosimilitud completa y reproduce la gráfica de $R(\theta; x)$.

Un intervalo de verosimilitud por sí solo resulta poco informativo y en consecuencia insuficiente para mostrar el cambio en la plausibilidad de los valores de θ en el intervalo; por tal razón en muchos casos se recomienda utilizar distintos intervalos de verosimilitud de nivel c e indicar en ellos al EMV $\hat{\theta}$ de θ . Esto permite identificar posibles asimetrías de la función de verosimilitud. En lo posible se recomienda graficar y analizar la función de verosimilitud relativa completa como se hará en el Capítulo 4.

A continuación se verá que es posible asociar una confianza deseada a los intervalos de verosimilitud, y que dicha confianza depende de la elección del nivel de plausibilidad c .

En particular, niveles de plausibilidad de $c = 0.036, 0.15, 0.25$ se encuentran asociados con niveles del 99, 95 y 90% de confianza (aproximadamente).

3.1.4 Intervalos de verosimilitud-confianza

Sea $x = (x_1, \dots, x_n)$ una muestra observada proveniente de una distribución de vaaid $X = (X_1, \dots, X_n)$ con función de probabilidad $P(\cdot; \theta)$, donde θ es un parámetro unidimensional fijo en un valor θ_0 . Entonces con base en esta muestra se puede construir un intervalo $[A, B]$ para el valor verdadero θ_0 . Ahora, si se construye nuevamente un intervalo $[A, B]$ pero con otra muestra del mismo experimento o fenómeno aleatorio de interés, casi seguramente se obtendrá un intervalo diferente al anterior. Esto ocurre debido a que los intervalos $[A, B]$ son variables aleatorias. Así, cada vez que se varíe la muestra, los intervalos $[A, B]$ algunas veces cubrirán el valor verdadero θ_0 y otras no.

La *probabilidad de cobertura* de un intervalo aleatorio $[A, B]$ es la probabilidad de que el intervalo $[A, B]$ cubra el valor verdadero θ_0 del parámetro,

$$PC(\theta_0) = P(A \leq \theta_0 \leq B; \theta = \theta_0).$$

En otras palabras, la probabilidad de cobertura $PC(\theta_0)$ expresa el porcentaje de veces que el intervalo $[A, B]$ cubre al valor verdadero θ_0 en un número muy grande de repeticiones de un experimento.

Obsérvese que en principio, la distribución de probabilidad de los extremos del intervalo A y B se puede calcular a partir de la distribución de la variable aleatoria X ; pero usualmente depende de θ_0 .

Kalbfleisch (1985, pág. 113) define un intervalo de confianza para θ de la siguiente forma. Un intervalo $[A, B]$ se llama *intervalo de confianza* para θ cuando su probabilidad de cobertura no depende de θ_0 . Es decir, cuando el valor de $PC(\theta_0)$ es el mismo para todo valor del parámetro θ_0 . En esta tesis se adoptará esta definición ya que es adecuada para los objetivos de este trabajo.

La probabilidad de cobertura de un $IV(c)$ se puede aproximar a través de la distribución de probabilidad de la estadística de la razón de verosimilitud para un θ fijo en θ_0 , $D_n = -2 \log R(\theta_0)$. A continuación se muestra esta metodología.

Para un valor particular $\theta = \theta_0$ se tiene la siguiente cadena de implicaciones,

$$\theta_0 \in IV(c) \Leftrightarrow R(\theta_0) \geq c \Leftrightarrow -2 \log R(\theta_0) \leq -2 \log(c).$$

De aquí que la probabilidad de cobertura del $IV(c)$ sea

$$\begin{aligned} PC(\theta_0) &= P[\theta_0 \in IV(c); \theta = \theta_0] \\ &= P[D_n \leq -2 \log(c); \theta = \theta_0]. \end{aligned} \quad (3.7)$$

Serfling (1980; pág. 155-156) prueba que, bajo ciertas condiciones, para todo $\theta_0 \in \Theta \subset \mathbb{R}$ se cumple que,

$$\lim_{n \rightarrow \infty} P(D_n \leq x; \theta = \theta_0) = P[\chi_{(1)}^2 \leq x], \forall x > 0.$$

Es decir que bajo algunas condiciones llamadas de regularidad, $D_n \equiv -2 \log R(\theta_0)$ converge en distribución a una Ji-cuadrada con un grado de libertad para todo $\theta_0 \in \Theta \subset \mathbb{R}$

Tomando en cuenta este resultado y la expresión (3.7) se tiene que la probabilidad de cobertura del $IV(c)$ es aproximadamente $P[\chi_{(1)}^2 \leq x]$, donde $x = -2 \log(c)$. Así, cuando $x = q_{(\alpha,1)}$, donde $q_{(\alpha,1)}$ es el cuantil $(1 - \alpha)$ de una distribución Ji-cuadrada con un grado de libertad, se tiene que el intervalo de verosimilitud de nivel c , con $c = \exp(-q_{(\alpha,1)}/2)$, heredará una probabilidad de cobertura aproximada del $100(1 - \alpha)\%$. Nótese que la probabilidad de cobertura del $IV(c)$ no depende de θ_0 de modo que el intervalo de verosimilitud es también un intervalo de confianza.

La Tabla 3.1 muestra los valores de c que se utilizan para calcular intervalos de verosimilitud-confianza con probabilidades de cobertura del 90, 95 y 99%. Por ejemplo, el valor 2.706 corresponde al cuantil 0.90 de una distribución Ji-cuadrada con un grado de libertad. Entonces un $IV(c)$ con $c = \exp(-2.706/2) = 0.258$ tiene una probabilidad

de cobertura del 90%. Análogamente, se pueden encontrar los valores de c que asocian una confianza del 95 y 99% a los intervalos de verosimilitud.

$(1 - \alpha)$	c	$q_{(\alpha,1)}$
0.90	0.258	2.706
0.95	0.146	3.841
0.99	0.036	6.635

Tabla 3.1: Confianza aproximada de intervalos de verosimilitud cuando θ es unidimensional.

3.1.5 Estimación en presencia de parámetros de estorbo: Verosimilitud perfil

Generalmente, cuando se modela un fenómeno aleatorio de interés, el modelo paramétrico elegido para ajustar a las observaciones de dicho fenómeno suele tener varios parámetros, de los cuales algunos pueden llegar a ser de gran interés y el resto considerarse como de poca trascendencia. De manera más formal se les denomina usualmente como parámetros de interés y parámetros de estorbo, respectivamente. Sin embargo, es importante hacer notar que un parámetro puede ser considerado en algunos casos como de estorbo y en otros casos no. Por ejemplo, si se tiene una muestra de variables aleatorias normales con media μ y varianza σ^2 y se desea estimar la media poblacional μ , entonces σ es un parámetro de estorbo. En contraste, si el parámetro de interés es σ , entonces ahora el parámetro de estorbo es μ .

Existen diferentes métodos para abordar el problema de estimación en presencia de parámetros de estorbo, véase Montoya (2008). En esta tesis se utilizará la verosimilitud maximizada o perfil ya que es un método estadístico general, sencillo y adecuado para los objetivos de este trabajo. A continuación se describe esta metodología.

Sea θ un vector conformado por un número finito de parámetros reales desconocidos, $\theta = (\theta_1, \dots, \theta_m) \in \Theta \subset \mathbb{R}^m$ con $m \geq 2$. Supóngase que haciendo un reacomodo en sus entradas puede ser escrito como $\theta = (\delta, \lambda)$, donde δ es de dimensión r y λ es de dimensión $m - r$, con $1 \leq r \leq m - 1$. Si δ es el parámetro de interés, entonces la función de verosimilitud maximizada o perfil para δ se define como:

$$L_P(\delta; x) = \max_{\lambda|\delta} L(\delta, \lambda; x) = L\left[\delta, \hat{\lambda}(\delta, x); x\right], \quad (3.8)$$

donde $\hat{\lambda}(\delta, x)$ es el estimador de máxima verosimilitud restringido (EMVR) de λ para un valor fijo de δ . El EMVR $\hat{\lambda}(\delta, x)$ es el valor de λ que cuenta con mayor plausibilidad para este valor fijo de δ dada la muestra observada $X = x$. De este modo, la verosimilitud perfil de δ , $L_P(\delta; x)$, se obtiene maximizando la función de verosimilitud global $L(\theta; x) = L(\delta, \lambda; x)$ sobre λ pero fijando δ .

La función de verosimilitud perfil relativa de δ se define como

$$R_P(\delta) = \frac{L_P(\delta; x)}{L_P(\hat{\delta}; x)} = \frac{L_P(\delta; x)}{L(\hat{\theta}; x)}, \quad (3.9)$$

donde $\hat{\theta} = (\hat{\delta}, \hat{\lambda})$ es el EMV de $\theta = (\delta, \lambda)$.

Notesé que para el caso particular donde $\theta = (\delta, \lambda)$ es un vector conformado por solo dos parámetros reales desconocidos, uno de interés (δ) y el otro de estorbo (λ), la gráfica de la función de verosimilitud, $L(\theta; x)$, es una superficie en \mathbb{R}^3 . A manera de ilustración, la Figura 3.1 muestra una superficie de verosimilitud relativa de δ y λ .

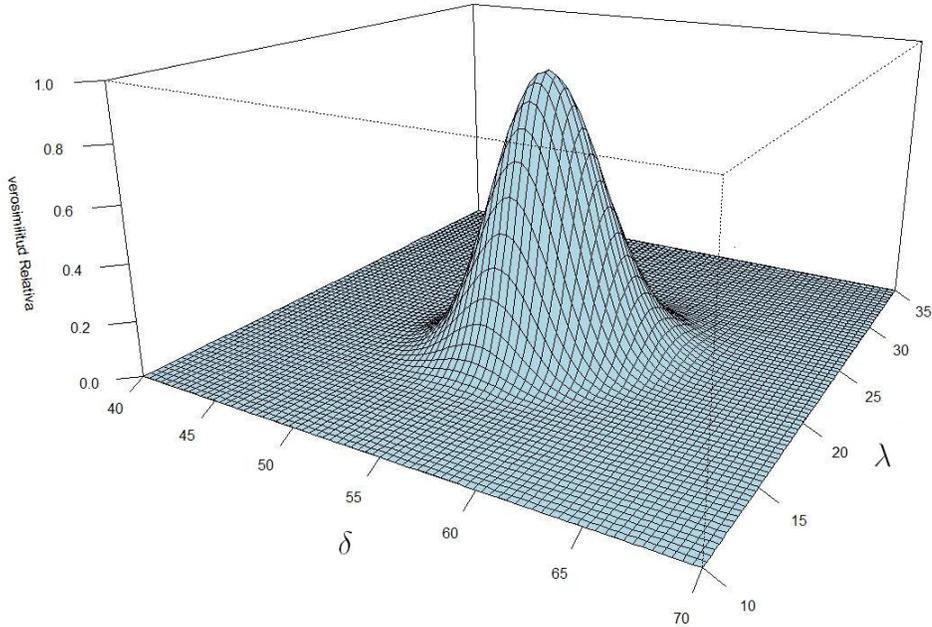


Figura 3.1: Superficie de verosimilitud relativa (δ, λ) .

Ahora, nótese que si se toma una posición en un punto distante sobre el eje designado a λ el parámetro de estorbo, entonces la silueta o el perfil observada de la verosimilitud global $L(\theta; x)$ es precisamente la función de verosimilitud maximizada de δ . Por ello, la verosimilitud maximizada es conocida también con el nombre de verosimilitud perfil.

Los intervalos de verosimilitud perfil de un parámetro de interés δ se definen como:

$$IVP(c) = \{\delta : R_P(\delta) \geq c\}, \quad \text{donde } 0 \leq c \leq 1.$$

Nótese que estos intervalos de verosimilitud son también intervalos de verosimilitud-confianza, véase Serfling (1980).

Cabe mencionar aquí que por la propiedad de invarianza de la verosimilitud, descrita en la Sección 3.1, la función de verosimilitud perfil es invariante frente a una reparametrización uno a uno del parámetro de estorbo. Es decir, la gráfica de la

función de verosimilitud perfil de un parámetro de interés δ no cambia cuando se reparametriza el parámetros de estorbo λ en términos de otro parámetro de estorbo ϕ , donde $\phi = \phi(\lambda)$ es una reparametrización uno a uno. Esto puede ser relevante puesto que ciertas reparametrizaciones pueden ser computacionalmente más simples de evaluar.

3.2 Métodos de valuación de modelos

Como ya se mencionó anteriormente, una parte muy importante dentro de un proceso de inferencia estadística es la valoración del modelo elegido para los datos. Existen varios métodos estadísticos para tal proposito. Aquí se describen algunos de ellos.

3.2.1 Métodos heurísticos

En esta sección se presentan métodos poco rigurosos pero sencillos y de fácil uso e interpretación para valorar modelos estadísticos.

Histogramas y densidades ajustadas

Después de analizar al contexto del fenómeno de interés y seleccionar un modelo para sus datos, una técnica sencilla para verificar que la elección puede ser adecuada es superponer en la misma gráfica el histograma de los datos y la densidad estimada del modelo (“el mejor” candidato) bajo los datos. Es importante mencionar que si el número de datos de la muestra es pequeño, por ejemplo $n = 5$, el concepto de histograma debe tomarse con mucha reserva, es decir, no se puede otorgar la misma significancia que cuando n fuera 100 o 500. Otro inconveniente con el que hay que tener cuidado es con la elección del número de clases o intervalos, porque si se consideran pocas, el histograma tendrá un aspecto demasiado burdo para ser informativo, y si se

consideran muchas, entonces el histograma tendrá un aspecto demasiado rugoso como para recoger aspectos relevantes.

Nótese que el histograma cumple aproximarse a la verdadera densidad de la variable aleatoria X usada para modelar el fenómeno, cualquiera que esta sea. No depende de suponer nada acerca de la naturaleza de la verdadera densidad de X .

Por otro lado, supóngase que X_1, \dots, X_n tienen densidad $f(x; \theta)$ y $\hat{\theta}$ es el EMV de θ . Sea $\hat{f}(x; \hat{\theta})$ la *densidad ajustada o estimada con los datos*. Si en efecto X_1, \dots, X_n tienen densidad $f(x; \theta)$, entonces la densidad ajustada \hat{f} converge a f . Pero si en realidad, la densidad de X es una, y para la estimación del parámetro vía el enfoque de verosimilitud se considero una alternativa diferente, entonces no necesariamente ocurrirá que la densidad ajustada se parezca a la densidad verdadera de X . Así, cuando la densidad estimada sea muy parecida a la forma del histograma entonces esto se considera un ajuste adecuado del modelo a los datos puesto que el histograma se aproxima al modelo verdadero. En contraste, cuando la densidad estimada y la forma del histograma no presenten una semejanza razonablemente buena, se determinará que el ajuste del modelo no es el adecuado para los datos.

El gráfico cuantil-cuantil

En estadística, una gráfica cuantil-cuantil es un método visual que permite verificar si un modelo estadístico es adecuado para la muestra observada. En lo que sigue, supóngase que se propone que la muestra observada x_1, \dots, x_n del fenómeno aleatorio bajo estudio proviene de una sucesión de variable aleatorias iid X_1, \dots, X_n con función de distribución $F(x; \theta)$, de la cual se desconoce θ . El objetivo es valorar que efectivamente $F(x; \theta)$ describe bien a los datos. La gráfica cuantil-cuantil se construye siguiendo los siguientes pasos:

1. Ordenar las observaciones de menor a mayor,

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Estas observaciones ordenadas son los llamados “cuantiles empíricos”.

2. Determinar los valores

$$p_i = \frac{i - 0.5}{n}, \quad i = 1, \dots, n.$$

Si por $Q_E(p)$ se denota al cuantil empírico de orden p ($0 < p < 1$) de las observaciones, se tiene que

$$x_{(i)} = Q_E(p_i), \quad i = 1, \dots, n.$$

3. Determinar los cuantiles de orden p_i , $i = 1, \dots, n$, de la distribución teórica representada por la función de distribución $F(x; \theta)$, es decir

$$Q_T(p_i; \theta) = F^{-1}(p_i; \theta), \quad i = 1, \dots, n.$$

Estos valores son los llamados “cuantiles teóricos”. Nótese que el valor de $Q_T(p_i; \theta)$ es desconocido puesto que el valor de θ no se conoce. Sin embargo, es posible aproximar $Q_T(p_i; \theta)$ con:

$$\hat{Q}_T(p_i) = Q_T(p_i; \hat{\theta}) = F^{-1}(p_i; \hat{\theta}), \quad i = 1, \dots, n,$$

donde $\hat{\theta}$ es un estimador puntual para θ . Por ejemplo, se puede tomar a $\hat{\theta}$ como el EMV de θ .

4. Representar en un plano cartesiano al conjunto de puntos ($\hat{Q}_T(p_i)$, $Q_E(p_i)$), $i = 1, \dots, n$ y a la recta $x = y$. Si estos puntos están muy cerca y serpenteando alrededor de la recta, es evidencia de que el modelo es adecuado. En contraste, cualquier patrón que de evidencia de que los puntos se alejan de la recta de 45° se interpretará como un mal ajuste del modelo a los datos.

Nube cuantil-cuantil

Esta técnica gráfica muestra si existe evidencia para rechazar o no rechazar al modelo elegido para los datos. El proceso de construcción de este gráfico se describe a continuación. Primero, suponer que se tiene un modelo adecuado para los datos y estimar sus parámetros. En esta etapa se pueden usar los estimadores de máxima verosimilitud. Luego, usar la densidad estimada para simular M muestras de tamaño n . Se recomienda tomar $M = 10,000$ y n igual que el tamaño de la muestra observada. Posteriormente, se estiman los cuantiles teóricos con la muestra observada y se calculan los cuantiles empíricos para cada muestra simulada. Nótese que los cuantiles teóricos estimados son siempre los mismos para cada muestra simulada. Para finalizar, se coloca en una misma figura la gráfica cuantil-cuantil para cada muestra simulada y se superponen junto con la de la muestra observada. Cuando la nube de puntos captura completamente a la recta de 45° entonces esto se considera como evidencia para no rechazar el modelo supuesto para los datos.

3.2.2 Uso de la verosimilitud perfil

Como se describió en la Sección 3.1.5, la verosimilitud perfil es una herramienta estadística que permite hacer inferencias sobre un parámetro de interés en presencia de otros de estorbo. Existen muchas situaciones en las que la distribución elegida para modelar un fenómeno de interés contiene a un parámetro que define a submodelos dentro de esa familia. Así, hacer inferencia sobre este parámetro vía la verosimilitud perfil (gráfica de los valores más plausibles del parámetro de interés e intervalos de verosimilitud) puede ayudar a elegir un modelo adecuado y parsimonioso.

Por ejemplo, un enfoque para seleccionar un modelo estadístico en la teoría de valores extremos es a través del uso de la distribución de valores extremos generalizada (DVEG) la cual cuenta con tres parámetros (a , b y c). El parámetro a es de localización, b es el de escala y c es el de forma (véase Coles, 2001). En este caso

el parámetro de interés es c y los de estorbo son a y b . Cuando $c < 0$ la DVEG se convierte matemáticamente en un modelo Weibull (de tres parámetros), cuando $c > 0$ se obtiene un modelo Fréchet (de tres parámetros) y si $c = 0$ se reduce a un modelo Gumbel (de dos parámetros). Así, una cuantificación de los valores plausibles de c puede ayudar a identificar al modelo adecuado para los datos dentro de esta familia de modelos. En este trabajo se propone utilizar la función de verosimilitud perfil de c para explorar gráficamente la preferencia de los datos por alguno de estos modelos. Si los valores plausibles de c se encuentran claramente a la derecha del valor $c = 0$ entonces se considera evidencia en contra del modelo Weibull y Gumbel. En este caso, se elige la distribución Fréchet para modelar el fenómeno bajo estudio. Si se encuentran claramente a la izquierda del valor $c = 0$ entonces se considera evidencia en contra del modelo Fréchet y Gumbel. En este caso, se elige la distribución Weibull. Por otro lado, si el valor $c = 0$ tiene alta plausibilidad entonces se puede considerar evidencia en contra del modelo Weibull y Fréchet. En este caso, se elige la distribución Gumbel. Para identificar el conjunto de valores plausibles para c se pueden usar los intervalos de verosimilitud perfil de niveles de plausibilidad 0.25, 0.15 y 0.036 (confianza aproximada de 90, 95 y 99% respectivamente).

3.2.3 Prueba de hipótesis

Una *prueba de hipótesis* consiste de examinar evidencia en forma de datos para dar lugar a una de dos resoluciones posibles: Rechazar H_0 a favor de H_1 , o no rechazar H_0 . Bajo este planteamiento, hay dos tipos de errores que se pueden cometer. El *error de Tipo I* se comete cuando se resuelve rechazar H_0 a favor de H_1 siendo que H_0 es “cierta” y el *error de Tipo II* se comete cuando se resuelve no rechazar H_0 cuando H_0 es “falsa”. Usualmente, en la teoría estadística se consideran las hipótesis de tal forma que el error de tipo I es más grave que el error de tipo II.

En general una prueba de hipótesis consta de dos ingredientes:

1. Una estadística $T = T(X_1, \dots, X_n)$ llamada la estadística de prueba.
2. Un subconjunto de valores posibles de T , llamado la región crítica, C , de la prueba.

El criterio a utilizar consiste en rechazar H_0 si y sólo si $T \in C$. Cuando $T \in C$ se dice que la prueba es significativa y cuando $T \notin C$ se dice que la prueba es no significativa. La región crítica no depende de la muestra x_1, \dots, x_n , lo que quiere decir que aún antes de tomar la muestra, la región crítica tiene existencia propia. Los datos intervienen para tomar o no la resolución de rechazar H_0 , lo cual se realiza con la región crítica, al comparar el valor de T con el conjunto C . Nótese que una prueba de hipótesis es de nivel α , con $0 \leq \alpha \leq 1$, si

$$\sup_{\theta \in \Theta_0} P_\theta [T \in C] \leq \alpha.$$

Por tal razón, muchas veces se denota a la región crítica C con el subíndice α para indicar que se trata de una prueba de nivel α .

En esta tesis se utiliza la estadística de prueba de la razón de verosimilitud definida como

$$D = -2 \log \left[\frac{L(\hat{\mu}, \hat{\sigma}; x)}{L(\hat{a}, \hat{b}, \hat{c}; x)} \right], \quad (3.10)$$

donde $L(\hat{\mu}, \hat{\sigma}; x)$ es la verosimilitud basada en el modelo Gumbel y $L(\hat{a}, \hat{b}, \hat{c}; x)$ es la verosimilitud basada en el modelo de la DVEG.

Esta estadística de prueba se distribuye como una χ^2 con grados de libertad igual a la diferencia de parámetros entre modelos (Casella y Berger 1990), en este caso es 1 grado de libertad. En las aplicaciones que se harán en el Capítulo 4 se elegirá una prueba de nivel $\alpha = 0.05$. Así, la región crítica quedará determinada por el percentil o cuantil 0.95 de una distribución χ^2 con 1 grado de libertad; es decir $C_\alpha = (3.84, \infty)$.

La teoría estadística sobre la que se basa el discurso para pruebas de hipótesis, recibe la denominación de teoría de Neyman-Pearson. Nótese que bajo este enfoque, dada una muestra observada y un nivel α el resultado de la prueba es binario, en el sentido de concluir “rechazar H_0 ” o “no rechazar H_0 ”. Esto puede ser criticable; por ejemplo, si la región crítica de una prueba de hipótesis fuera el conjunto $C_\alpha = (3.84, \infty)$ y datos en dos situaciones diferentes dieran lugar a valores de la estadística de prueba $D_1 = 4.23$ y $D_2 = 9.17$. La actitud de Neyman-Pearson diría simplemente, en ambos casos, “rechazar H_0 ”, siendo que es intuitivamente claro que ambas situaciones son diferentes en alguna cualidad. En el segundo caso, se rechaza con mayor fuerza que en el primero, y al decir sólo “rechazar H_0 ” no involucramos esta fuerza de la evidencia en contra de H_0 . Un enfoque que considera la evidencia en contra de una hipótesis es descrito a continuación.

3.2.4 Pruebas de significancia: *p-valor*

El concepto de *p-valor* tiene por objeto cuantificar la fuerza con la que se rechaza una hipótesis nula H_0 . Se describe a través de una probabilidad. Tiene la interpretación de ser la probabilidad de haber observado un valor “más extremo” de una estadística de prueba $T(X)$ que ya se observó $t_0(x)$, o bien, la probabilidad de haber rechazado H_0 sólo por azar. La definición matemática del *p-valor* es

$$p\text{-valor} = P(x) = \sup_{\theta \in \Theta_0} P_\theta [T(X) \geq t_0(x)],$$

(Mood et al. 1985).

De esta forma, un *p-valor* “grande” denota que la evidencia en contra de H_0 es débil y un *p-valor* “chico” denota que los datos contienen mucha evidencia en contra de H_0 . En este sentido de *p-valores*, se puede no hablar de pruebas de hipótesis, sino de pruebas de significancia, donde la cuantificación del concepto abstracto de significancia es el *p-valor*.

En este trabajo de tesis para calcular el *p-valor* se considerará la estadística de prueba (3.10) que se distribuye como una χ_1^2 . Así, el *p-valor* se puede calcular a través de la siguiente expresión:

$$p\text{-valor} = P [\chi_1^2 > D_0] = 1 - P [\chi_1^2 \leq D_0], \quad (3.11)$$

donde D_0 es el valor de la estadística (3.10) bajo los datos observados.

Estas técnicas estadísticas son muy útiles para valorar al modelo que será utilizado para hacer las inferencias. En el siguiente capítulo se aplican los resultados descritos en los Capítulos 2 y 3 a un problema real e importante.

Capítulo 4

Caso de estudio: Precipitación máxima anual en Hermosillo, Sonora

En este capítulo se hace un análisis estadístico de los niveles de precipitación máximos anuales observados entre los años 1965 y 2009 en Hermosillo, Sonora. Esto con el objetivo de mostrar un proceso completo de inferencia estadística (modelación, valoración del modelo, inferencia sobre parámetros de interés en el contexto del problema e interpretación de los resultados en el contexto del problema) en un problema relevante de la sociedad. En particular, se hace inferencia sobre dos parámetros que pueden ser de gran interés para la gestión de riesgos, el primero es el denominado nivel de retorno, que es frecuentemente utilizado en muchas ramas de la ingeniería, y el segundo es un parámetro propuesto en este trabajo que se llamará parámetro de vulnerabilidad a inundaciones. Además, se aplican las diferentes técnicas de valuación de modelos, descritas en la Sección 3.2, con el objetivo de justificar la elección del modelo con base en los datos observados.

4.1 Descripción del problema

Como se mencionó en el Capítulo 1, una forma de analizar la vulnerabilidad a inundaciones es a través del estudio del NPMA en un periodo de tiempo razonable. Los datos de niveles de precipitación máximos anuales (1965-2009) de Hermosillo, Sonora, que serán utilizados aquí, fueron proporcionados por la CONAGUA y provienen de dos estaciones meteorológicas ubicadas en esta ciudad, la Estación Observatorio y la Estación Norte (ver Figura 4.1).

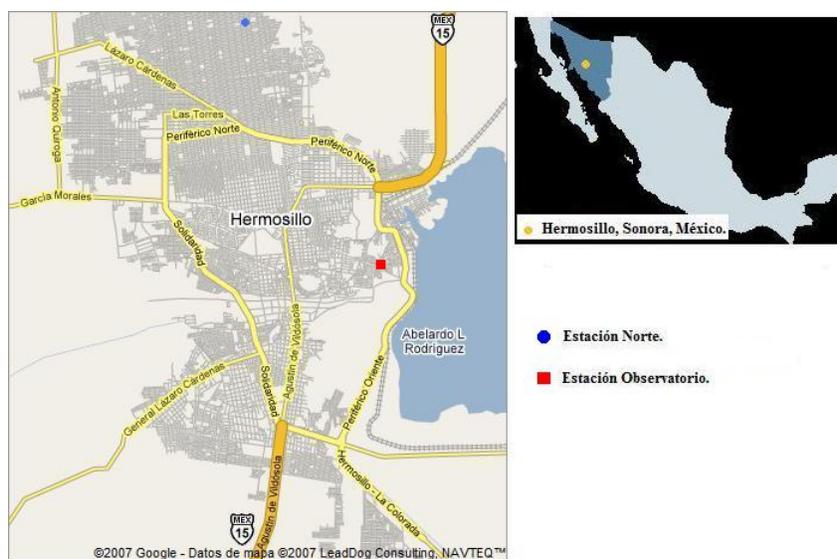


Figura 4.1: Estaciones meteorológicas de la CONAGUA en Hermosillo, Sonora.

Las estaciones Observatorio y Norte se encuentran ubicadas geográficamente cerca. Además, las condiciones meteorológicas entre ambas estaciones no cambian sustancialmente. En este sentido, no es inadecuado “combinar” la información de las dos estaciones y determinar una sola muestra representativa de la región. La metodología para obtener esta muestra es la siguiente. Se toma el valor máximo por año entre las dos estaciones (máximo de la región) y en los años donde sólo se tiene un registro se toma a ese valor como el NPMA de la zona. En la Figura 4.2 se muestran los datos que representarán a la región.

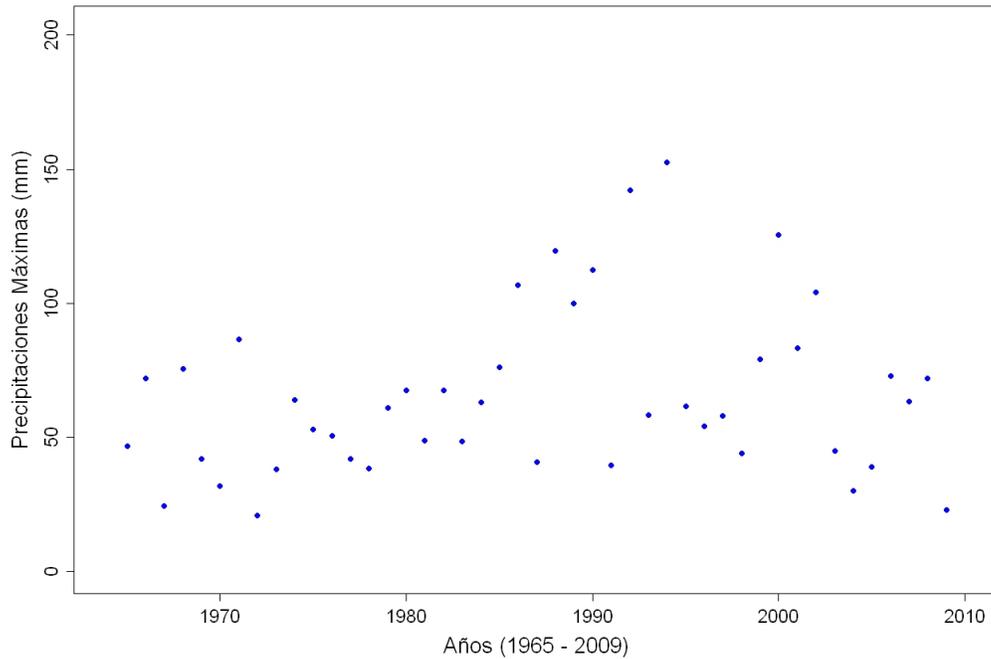


Figura 4.2: Registros del NPMA para Hermosillo, Sonora, México.

Ahora, dada la muestra de niveles máximos de precipitación en la región, el problema estadístico es hacer inferencia sobre el parámetro llamado nivel de retorno y sobre otro parámetro propuesto en esta tesis denominado parámetro de vulnerabilidad a inundaciones. A continuación se describen estos parámetros.

4.2 Parámetros de interés

En esta sección se describen dos parámetros relevantes para la gestión de riesgos. El primero llamado nivel de retorno, un parámetro ya muy usual y ampliamente utilizado en las distintas ramas de la ingeniería y el segundo, propuesto en este trabajo, denominado parámetro de vulnerabilidad a inundaciones.

4.2.1 Nivel de retorno

En muchas ramas de la ingeniería es de interés conocer cuál es el valor que en promedio se excede una vez cada determinado tiempo. Para tal propósito, se considerarán observaciones independientes de variables aleatorias con función de distribución F conocida, tomadas en periodos de tiempo iguales.

En general, sea $\{X_i\}_{i \geq 1}$ una sucesión de variables aleatorias (iid) con función de distribución F y w un valor dado. Considérese la sucesión $\{1_{\{X_i > w\}}\}_{i \geq 1}$ de variables aleatorias (iid) Bernoulli que toman el valor 1, “éxito”, si $X_i > w$ y 0 en otro caso, con probabilidad $p = 1 - F(w)$ y $1 - p$, respectivamente. Entonces, el instante del primer éxito

$$Y(w) = \min \{i \geq 1 : X_i > w\},$$

es decir, el instante de la primera excedencia del valor w , es una variable aleatoria geométrica con función de probabilidad

$$P[Y(w) = k] = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

Por lo tanto

$$E[Y(w)] = \frac{1}{p}.$$

Al valor w se le llama nivel de retorno con periodo de retorno $1/p$ para los eventos $\{X_i > w\}_{i \geq 1}$. Así, si se desea encontrar el nivel de retorno w correspondiente a un periodo de 20 años, entonces se debe encontrar el valor de w tal que $E[Y(w)] = 20$; es decir, encontrar w tal que $p = 1/20 = 0.05$. Ahora, como $p = 1 - F(w)$ entonces el problema consiste en resolver la ecuación $F(w) = 1 - 0.05 = 0.95$ para w . Por lo tanto, w es el cuantil 0.95 de la distribución F , $Q_{0.95}$. Así, el *nivel de retorno* de un periodo fijo de t años se define en general como el cuantil $1 - \alpha(t)$ de la distribución F , $Q_{1-\alpha(t)}$, donde $\alpha(t) = 1/t$.

Otro parámetro que se considera pueda ser de gran utilidad para la gestión de riesgos o en general para alguna rama de la ingeniería se describe a continuación.

4.2.2 Vulnerabilidad a inundaciones

Supóngase que se tiene una estructura la cual tiene como principal factor de vulnerabilidad la precipitación de lluvia. En particular, se supone que la estructura colapsará si el nivel de precipitación sobrepasa un umbral v fijo y conocido. Nótese que el valor crítico v depende de la estructura en cuestión.

Por otro lado, supóngase que el NPMA se puede modelar con una variable aleatoria que tiene función de distribución F . El *parámetro de vulnerabilidad a inundaciones* se define como

$$\theta = P[X > v; \lambda], \quad (4.1)$$

donde v es un valor dado y λ es el vector de parámetros asociados a F . Así, el parámetro de vulnerabilidad θ dado en (4.1) es la probabilidad de que la estructura de nivel crítico v colapse por precipitación. Para mayores detalles sobre el uso de este parámetro en el análisis de vulnerabilidad de sistemas estructurales véase Murray y Grubestic (2007).

El parámetro llamado nivel de retorno, descrito en la sección anterior, y el parámetro de vulnerabilidad dado en (4.1) serán estimados más adelante, ya que para ello se requiere primero determinar cual es el modelo F adecuado para los datos.

4.3 Modelación

El fenómeno aleatorio que se desea modelar es el NPMA en Hermosillo. Es decir, se tiene interés en una distribución de probabilidad para el valor más alto de la precipitación de lluvia durante un periodo de un año en esta localidad. Así, por el contexto

del problema, es razonable utilizar alguna de las distribuciones de la teoría de valores extremos, descritas en el Capítulo 2. Considerando lo anterior, aquí se propone usar alguno de los submodelos (Gumbel, Weibull o Fréchet) asociados a la DVEG, para modelar el NPMA en Hermosillo, Sonora.

Cabe recordar que el modelo Gumbel es el miembro más simple de la familia de distribuciones de valores extremos generalizada en el sentido de que este modelo cuenta con sólo dos parámetros (los modelos Weibull y Fréchet tienen tres parámetros). Por tal razón, en la siguiente sección se explorará si en particular el modelo Gumbel es adecuado para modelar el NPMA en Hermosillo, Sonora.

4.4 Valoración de modelos: Weibull, Gumbel y Fréchet

En el Capítulo 2 se presentaron los modelos Weibull, Gumbel y Fréchet como submodelos de la DVEG la cual depende de tres parámetros, a , b y c . Se mencionó además que el uso de alguna de estas distribuciones para modelar datos de extremos depende del valor de c . El valor de $c < 0$ corresponde al modelo Weibull, el valor de $c > 0$ corresponde al modelo Fréchet y el valor de “ $c = 0$ ” corresponde a un modelo Gumbel. A continuación se usaran los métodos de valuación de modelos presentados en la Sección 3.2, con el objetivo de valorar el ajuste de estos modelos a los datos mostrados en la Figura 4.2. Además se usará la teoría de estimación vía el enfoque de verosimilitud, descrita en la Sección 3.1, para la estimación de parámetros.

La función de verosimilitud del vector de parámetros (a, b, c) de la DVEG se puede

escribir de la siguiente forma:

$$\begin{aligned}
 L(a, b, c; x) &\propto \prod_{i=1}^n f(x_i; a, b, c) \\
 &= (b)^{-n} \prod_{i=1}^n \left[1 + c \left(\frac{x_i - a}{b} \right) \right]^{-(1+\frac{1}{c})} \\
 &\quad \exp \left\{ - \sum_{i=1}^n \left[1 + c \left(\frac{x_i - a}{b} \right) \right]^{-\frac{1}{c}} \right\},
 \end{aligned}$$

donde x_1, \dots, x_n son los datos observados. Los valores que maximizan esta función son: $\hat{a} = 50.6732$, $\hat{b} = 22.3435$ y $\hat{c} = 0.0793$. Nótese que el EMV del parámetro c , $\hat{c} = 0.0793$, es positivo; esto indica que el modelo Fréchet es adecuado para los datos. Sin embargo, \hat{c} también está cercano a cero; lo cual sugiere que un modelo Gumbel también pudiera ser razonable a la luz de los datos. En contraste, el modelo Weibull no es adecuado bajo este criterio.

Una estimador puntual no proporciona información acerca de la incertidumbre en la estimación. Para mostrar que valores de un parámetro son razonables o contradichos por los datos, se recomienda usar la función de verosimilitud perfil. En la Figura 4.3 se muestra la verosimilitud perfil del parámetro c y un intervalo de verosimilitud-confianza del 95%. Se observa que valores de c cercanos a cero y un poco mayores a este valor tienen alta plausibilidad. Además, el intervalo de verosimilitud-confianza del 95% para c (nivel de verosimilitud del 0.15), mostrado en la Figura 4.3, captura el valor $c = 0$ asociado a un modelo Gumbel. Por otro lado, valores muy negativos de c son poco plausibles. Así, el modelo Weibull es poco creíble a la luz de la muestra observada. En contraste, los modelos Gumbel y Fréchet no pueden ser descartados para ajustar este juego de datos.

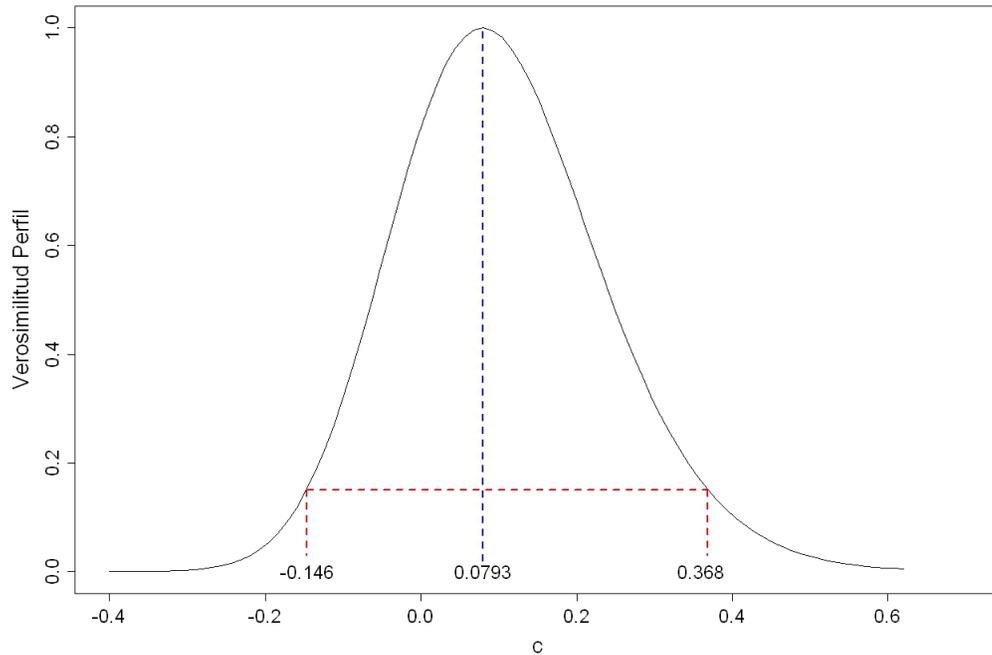


Figura 4.3: Verosimilitud perfil del parámetro c de la DVEG bajo los datos de estudios mostrados en la Figura 4.2. Se señala un intervalo de verosimilitud-confianza de 95% (nivel de plausibilidad de 0.15).

La Figura 4.4 muestra las densidades Gumbel y Fréchet ajustadas al histograma de frecuencias relativa de los datos. Los parámetros de estos modelos fueron estimados mediante el método de máxima verosimilitud.

(i) Modelo Gumbel: $\hat{\mu} = 51.64$ y $\hat{\sigma} = 23.06$.

(ii) Modelo Fréchet: $\hat{\mu} = -230.15$, $\hat{\sigma} = 280.82$ y $\hat{\beta} = 12.56$.

Se observa que los dos modelos se ajustan razonablemente bien al histograma.

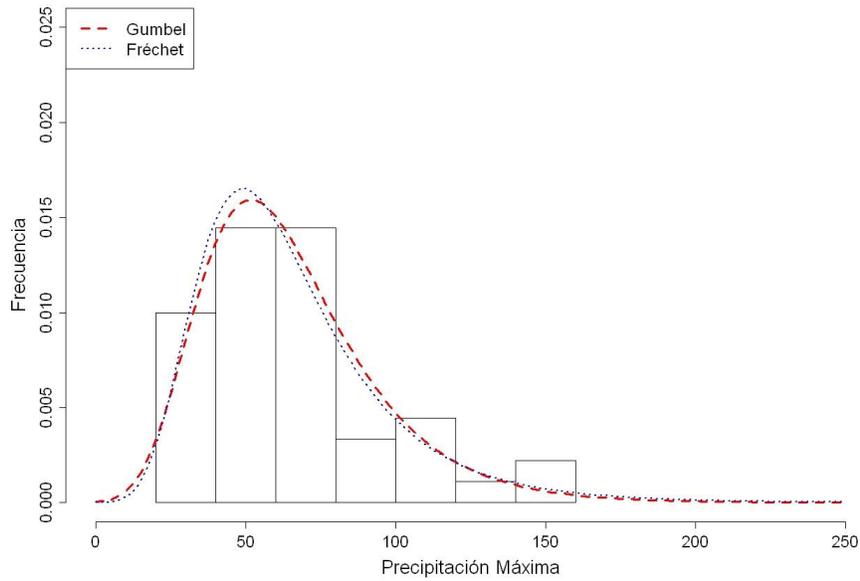


Figura 4.4: Histograma con la densidad ajustada del modelo Gumbel y la densidad ajustada del modelo Fréchet.

Las Figuras 4.5 y 4.6 corresponden a las gráficas cuantil-cuantil bajo el “mejor” modelo Gumbel y el “mejor” modelo Fréchet, respectivamente. Es decir, los parámetros de los modelo fueron reemplazados por sus correspondientes estimadores de máxima verosimilitud.

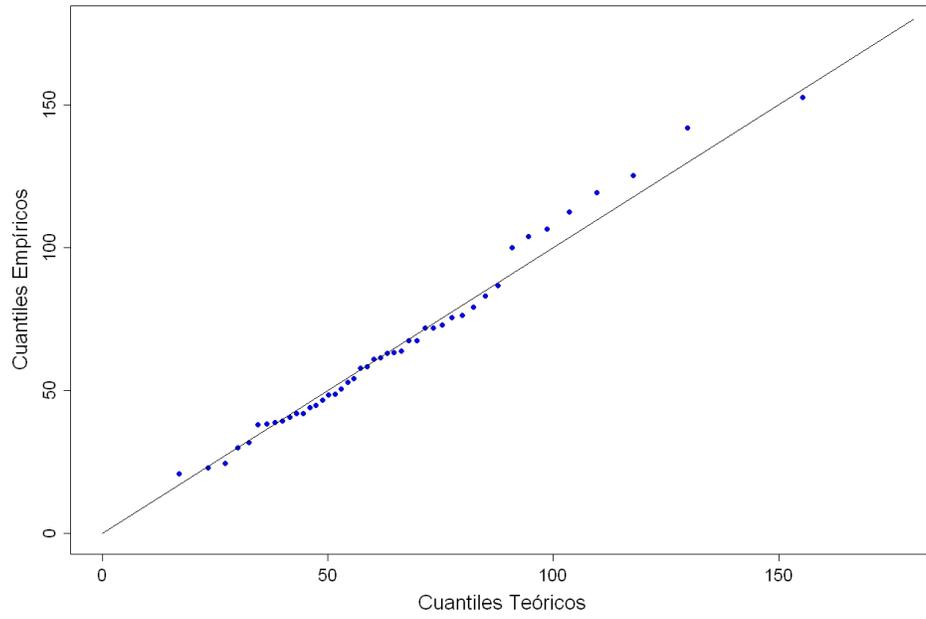


Figura 4.5: Gráfica cuantil-cuantil para el “mejor” modelo Gumbel.

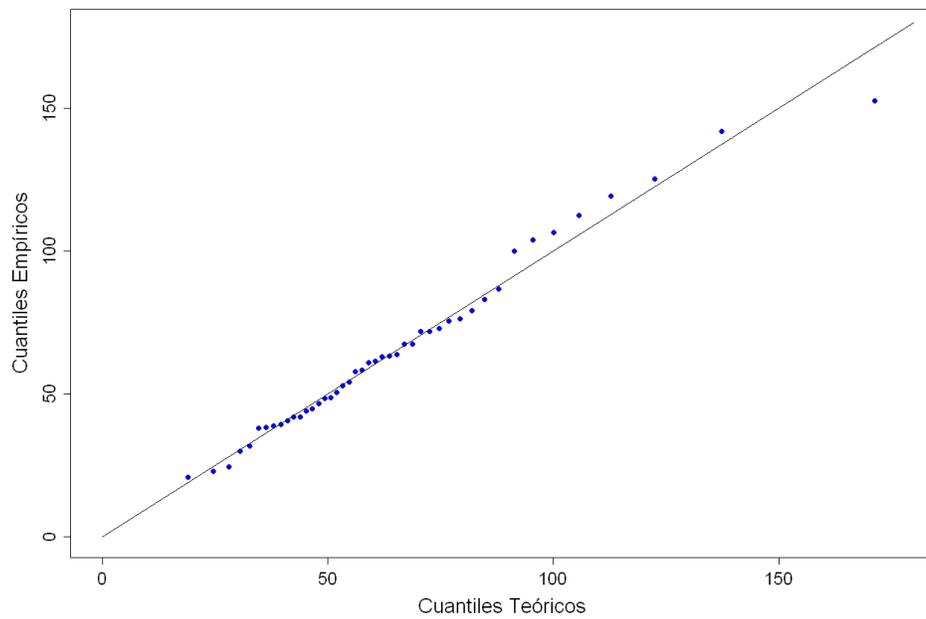


Figura 4.6: Gráfica cuantil-cuantil para el “mejor” modelo Fréchet.

En las Figuras 4.5 y 4.6 se observa que los puntos serpentean alrededor de la recta $y = x$. Nótese que a pesar de que el modelo Gumbel tiene un parámetro menos que el modelo Fréchet, la Figura 4.5 muestra un ajuste razonablemente bueno del modelo a los datos.

Otra herramienta ilustrativa que proporciona evidencia acerca del ajuste adecuado de un modelo es la nube cuantil-cuantil, descrita en la Sección 3.2. La Figura 4.7 presenta esta gráfica para el “mejor” modelo Gumbel.

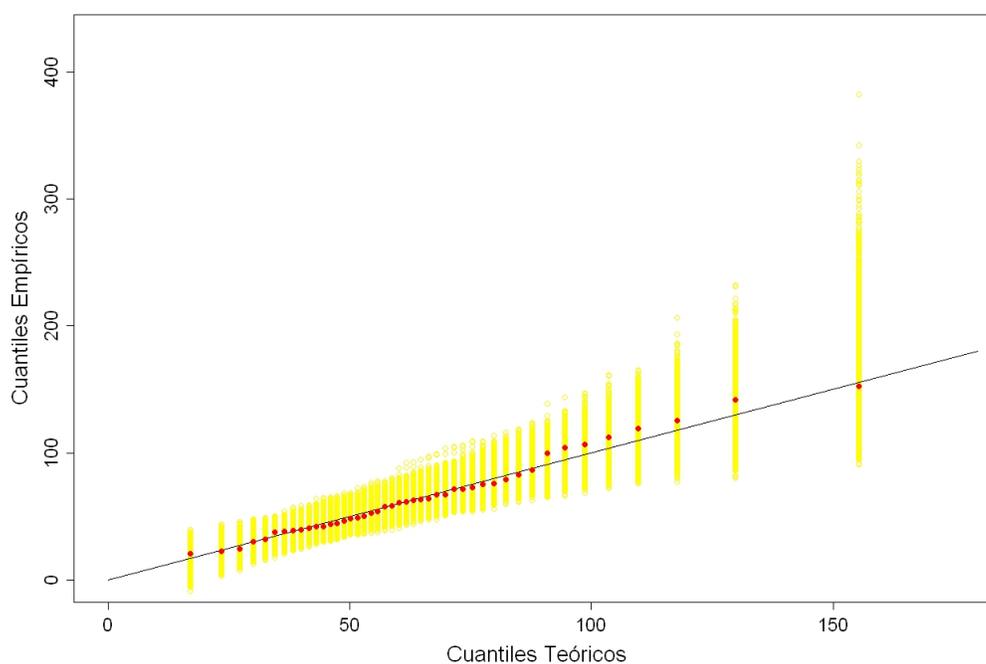


Figura 4.7: Nube cuantil-cuantil para el “mejor” modelo Gumbel.

En la Figura 4.7 puede observarse la variabilidad de los cuantiles empíricos con respecto a los cuantiles teóricos determinados bajo el “mejor” modelo Gumbel. Se observa una alta variabilidad para cuantiles altos. Obsérvese que la recta de 45° se encuentra totalmente contenida en la nube de puntos; esto indica que el modelo evaluado (en este caso el Gumbel) puede ser adecuado para los datos.

Además de herramientas visuales para la valuación de modelos existen también herramientas teóricas. A continuación se aplica la prueba de hipótesis y la prueba de significancia del *p-valor*, descritas en el capítulo anterior, tanto para determinar si se rechaza o no el modelo Gumbel como para conocer la evidencia en contra de éste, respectivamente.

Prueba de hipótesis

Se realizó una prueba de hipótesis con las características descritas en la Sección 3.2. La hipótesis nula H_0 es que el “modelo Gumbel es adecuado”. El valor de la estadística de prueba es

$$D_0 = -2 \log \left[\frac{L(\hat{\mu}, \hat{\sigma}; x)}{L(\hat{a}, \hat{b}, \hat{c}; x)} \right] = 0.40, \quad (4.2)$$

y la región crítica o área de rechazo de nivel $\alpha = 0.05$ es $C_\alpha = (3.84, \infty)$. Así, no se rechaza la hipótesis nula H_0 . Es decir, no se rechaza el modelo Gumbel con una confianza del 95%.

Pruebas de significancia: *p-valor*

Nótese que al no rechazar H_0 se espera que el *p-valor* sea mayor que $\alpha = 0.05$. Sin embargo, se procederá a calcularlo para fines de ejemplificar esta metodología.

Usando la estadística dada en (4.2) y la expresión (3.11), se obtiene el siguiente resultado:

$$\begin{aligned} p - valor &= P[\chi_1^2 > D_0] \\ &= 1 - P[\chi_1^2 \leq 0.40] = 0.52. \end{aligned}$$

Así, como se había previsto, el *p-valor* encontrado no refleja evidencia contundente en contra del modelo Gumbel.

Considerando cada evidencia proporcionada por todos los métodos de valuación de modelos aplicados aquí, se decide que el modelo Gumbel de dos parámetros es adecuado y parsimonioso para modelar el NPMA de la ciudad de Hermosillo, Sonora.

4.5 Inferencias sobre parámetros de interés usando la verosimilitud perfil

En esta sección se usará el modelo Gumbel y el enfoque de verosimilitud perfil para realizar inferencias sobre el parámetro llamado nivel de retorno y sobre el parámetro de vulnerabilidad a inundaciones, ambos descritos en la Sección 4.2. Es importante señalar que en cada caso, se puede reparametrizar de forma muy simple el modelo Gumbel en términos del correspondiente parámetro de interés. Esto permite escribir la verosimilitud en función del parámetro de interés y otro de estorbo. Así, el enfoque de verosimilitud perfil descrito en la Sección 3.1 se aplica de manera directa. A continuación se describe el proceso de reparametrización de modelo Gumbel para cada caso.

El nivel de retorno de un periodo fijo de t años se define como el cuantil $1 - \alpha$, $Q_{1-\alpha}$, de la distribución Gumbel,

$$Q_{1-\alpha} = \mu - \sigma \{ \log [- \log (1 - \alpha)] \},$$

donde $\alpha = 1/t$. Entonces, el parámetro de localización μ de la distribución Gumbel se puede expresar como función de los parámetros $Q_{1-\alpha}$ y σ de la siguiente forma:

$$\mu = \mu(Q_{1-\alpha}, \sigma) = Q_{1-\alpha} + \sigma \{ \log [- \log (1 - \alpha)] \}. \quad (4.3)$$

Así, el modelo Gumbel queda escrito en términos de $Q_{1-\alpha}$ y σ , donde $Q_{1-\alpha}$ es el parámetro de interés y σ es el parámetro de estorbo. Considerando esta reparametrización del modelo, se pueden hacer inferencias sobre $Q_{1-\alpha}$ usando la función de

verosimilitud perfil de $Q_{1-\alpha}$ calculada a partir de la función de verosimilitud de $Q_{1-\alpha}$ y σ ,

$$L(Q_{1-\alpha}, \sigma; x) \propto \prod_{i=1}^n \frac{1}{\sigma} \exp \left\{ - \exp \left[\frac{\mu(Q_{1-\alpha}, \sigma) - x_i}{\sigma} \right] + \frac{\mu(Q_{1-\alpha}, \sigma) - x_i}{\sigma} \right\},$$

donde $\mu(Q_{1-\alpha}, \sigma)$ esta dado en (4.3), x_1, \dots, x_n son los datos observados y $\alpha = 1/t$ (t es un periodo fijo de años).

Para estimar el parámetro de vulnerabilidad a inundaciones se usará una reparametrización del modelo Gumbel diferente a la anterior. El parámetro de vulnerabilidad $\theta = P[X > v; \mu, \sigma]$, bajo el supuesto de que la variable aleatoria X tiene una distribución Gumbel, se puede expresar como:

$$\theta = 1 - \exp \left\{ - \exp \left[- \left(\frac{v - \mu}{\sigma} \right) \right] \right\},$$

donde v es una cantidad fija y conocida. Entonces, el parámetro de localización μ de la distribución Gumbel se puede expresar ahora como función de los parámetros θ y σ de la siguiente forma:

$$\mu = \mu(\theta, \sigma) = v + \sigma \log [-\log (1 - \theta)]. \quad (4.4)$$

Así, el modelo Gumbel queda reparametrizado ahora en términos de θ y σ , donde θ es el parámetro de interés y σ es el parámetro de estorbo. Considerando esta nueva reparametrización del modelo, se pueden hacer inferencias sobre θ usando la función de verosimilitud perfil de θ calculada a partir de la función de verosimilitud de θ y σ ,

$$L(\theta, \sigma; x) \propto \prod_{i=1}^n \frac{1}{\sigma} \exp \left\{ - \exp \left[\frac{\mu(\theta, \sigma) - x_i}{\sigma} \right] + \frac{\mu(\theta, \sigma) - x_i}{\sigma} \right\},$$

donde $\mu(\theta, \sigma)$ esta dado en (4.4) y x_1, \dots, x_n son los datos observados.

En las siguientes secciones se muestran estimaciones de los parámetros de interés, $Q_{1-\alpha}$ y θ , con base en los niveles de precipitación máximos anuales observados en Hermosillo, Sonora, presentados en la Figura 4.2.

4.5.1 Nivel de retorno

En general, en teoría de valores extremos es de interés conocer niveles de retorno asociados a periodos de 20 y 100 años, respectivamente. Sin embargo, los periodos de retorno pueden variar dependiendo de la conveniencia de quien lo utiliza. La Figura 4.8 muestra la gráfica de la función de verosimilitud perfil relativa del nivel de retorno $Q_{0.95}$, asociado a un periodo de retorno de $t = 20$ años. El valor más plausible para el nivel de retorno $Q_{0.95}$ es $\hat{Q}_{0.95} = 120.15$ mm y el intervalo de verosimilitud del 95% de confianza para $Q_{0.95}$ es $IV(0.15) = [103.53, 142.81]$. Tanto el EMV de $Q_{0.95}$ como el intervalo de verosimilitud anterior son señalados en la Figura 4.8.

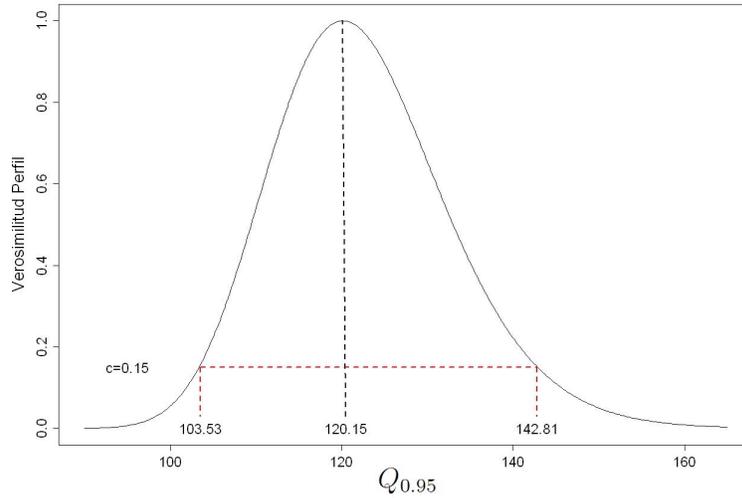


Figura 4.8: Gráfica de la verosimilitud perfil relativa del nivel de retorno $Q_{0.95}$. Se marca el EMV de $Q_{0.95}$ y el $IV(c)$ del 95% de confianza (nivel de plausibilidad $c = 0.15$).

Análogamente, la Figura 4.9 muestra la gráfica de la función de verosimilitud perfil relativa del nivel de retorno $Q_{0.99}$, asociado a un periodo de retorno de $t = 100$ años. En este caso el EMV de $Q_{0.99}$ es $\hat{Q}_{0.99} = 157.75$ mm y el intervalo de verosimilitud del 95% de confianza es $IV(0.15) = [133.88, 190.57]$.

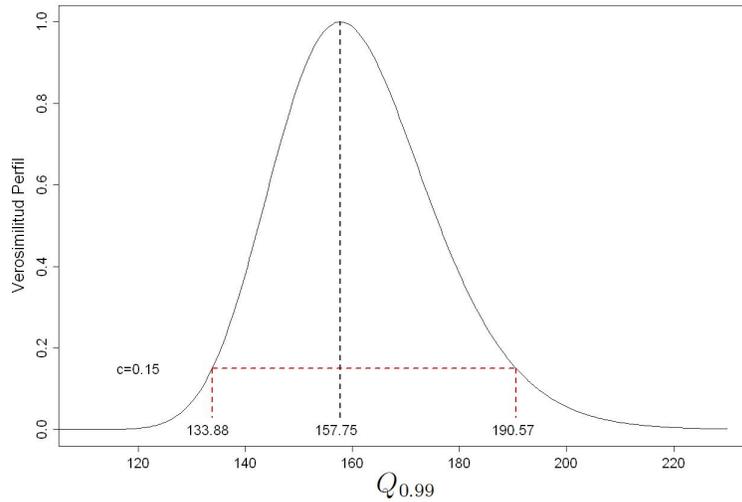


Figura 4.9: Gráfica de la verosimilitud perfil relativa del nivel de retorno $Q_{0.99}$. Se marca el EMV de $Q_{0.99}$ y el $IV(c)$ del 95% de confianza (nivel de plausibilidad $c = 0.15$).

Nótese que el nivel de retorno estimado para un periodo de 100 años es mayor que el de 20 años ($157.75 > 120.15$). Esto se debe fundamentalmente a que la verosimilitud usa toda la estructura matemática del modelo, y en particular que el cuantil asociado al periodo de retorno de 100 años es teóricamente más grande que el cuantil correspondiente a un periodo de retorno de 20 años.

En la Tabla 4.1 y 4.2 se presenta el resumen de las inferencias en términos de verosimilitud para los niveles de retorno asociados a 20 y 100 años, respectivamente. Nótese que A y B son los extremos del intervalo de verosimilitud-confianza para el parámetro de interés.

NIVEL	A	$Q_{0.95}$	B
c=0.036	98.83	120.15	151.84
c=0.15	103.53	120.15	142.81
c=0.25	105.60	120.15	139.06

Tabla 4.1: Resumen de inferencias para el nivel de retorno asociado a un periodo de 20 años.

NIVEL	A	$Q_{0.99}$	B
c=0.036	127.69	157.75	203.60
c=0.15	133.88	157.75	190.57
c=0.25	137.14	157.75	185.03

Tabla 4.2: Resumen de inferencias para el nivel de retorno asociado a un periodo de 100 años.

4.5.2 Vulnerabilidad a inundaciones

Para estimar el parámetro de vulnerabilidad a inundaciones, $\theta = P[\text{NPMA} > v]$, es necesario fijar el valor crítico v . Recuérdese que este valor no es estándar y depende de la situación particular que se desee evaluar (véase Sección 4.2.2). Por ejemplo, v puede representar el nivel máximo de precipitación para que el sistema de drenaje en una zona geográfica específica colapse o el nivel máximo de precipitación requerido para inundar un camino y se vuelva intransitable, etc.

Para efectos de ejemplificar las estimaciones de verosimilitud perfil sobre θ , aquí se tomará el valor $v = 80$ mm. El valor más creíble para θ bajo los datos de estudio fue $\hat{\theta} = 0.2535$. El intervalo de verosimilitud confianza del 95% para θ es $IV(0.15) = [0.1614, 0.3675]$, véase la Figura 4.10.

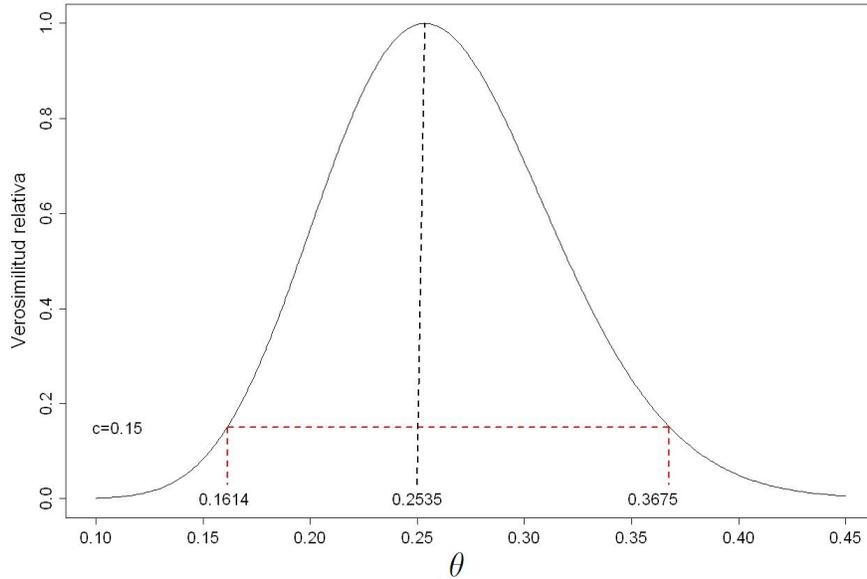


Figura 4.10: Gráfica de la verosimilitud perfil relativa para el parámetro de vulnerabilidad a inundaciones $\theta = P[\text{NPMA} > v]$, donde $v = 80$ mm. Se marca el EMV de θ y el $IV(c)$ del 95% de confianza (nivel de plausibilidad $c = 0.15$).

En la Tabla 4.3 se presenta el resumen de las inferencias en términos de verosimilitud para el parámetro de vulnerabilidad a inundaciones θ . Nótese que A y B son los extremos del intervalo de verosimilitud-confianza para θ .

NIVEL	A	$\hat{\theta}$	B
$c=0.036$	0.1368	0.2535	0.4078
$c=0.15$	0.1614	0.2535	0.3675
$c=0.25$	0.1736	0.2535	0.3500

Tabla 4.3: Resumen de inferencias para el parámetro de vulnerabilidad a inundaciones θ .

4.6 Interpretación de resultados

En la sección anterior se realizaron inferencias vía el enfoque de verosimilitud perfil sobre el nivel de retorno y el parámetro de vulnerabilidad a inundaciones. En esta sección se hace una interpretación de estas inferencias en el contexto del problema.

4.6.1 Nivel de retorno

El EMV del nivel de retorno asociado a un periodo de retorno de 20 años fue $\hat{Q}_{0.95} = 120.15$. Es decir, se estima que en 20 años se espera obtener un NPMA mayor que 120.15 mm. Por otro lado, el intervalo de verosimilitud-confianza del 95% para $Q_{0.95}$ fue $IV(0.15) = [103.53, 142.81]$. Es decir, con una confianza del 95%, el NPMA que será sobrepasado una vez en 20 años se encuentra entre 103.53 mm y 142.81 mm. En la Figura 4.11 se presentan estas estimaciones junto con los niveles de precipitación máximos anuales observados entre los años 1965 y 2009 en Hermosillo, Sonora. Además, se señala el periodo 2009-2029 (con líneas verticales) correspondiente a 20 años y se muestra el NPMA observado en el 2010, no considerado en el proceso de estimación. Nótese que este valor no sobrepasó el umbral $\hat{Q}_{0.95} = 120.15$ mm. De hecho, no sobrepasó el límite inferior, 103.53 mm, correspondiente al intervalo de confianza del 95% para $Q_{0.95}$, pintado en la Figura 4.11. Nótese que todo este análisis se puede hacer para cualquier nivel de retorno asociado a un periodo de retorno de t años.

En la práctica, la toma de decisiones con base en estimaciones del nivel de retorno seguramente dependerá de que tan precavido se quiera llegar a ser con respecto a las consecuencias de una mala decisión. En este sentido, para el “muy precavido” se le puede recomendar usar el límite superior de un intervalo de confianza. En contraste, para el “poco precavido” se le puede recomendar usar el límite inferior de un intervalo de confianza. Emplear el valor más creíble a la luz de los datos (EMV) es una opción intermedia.

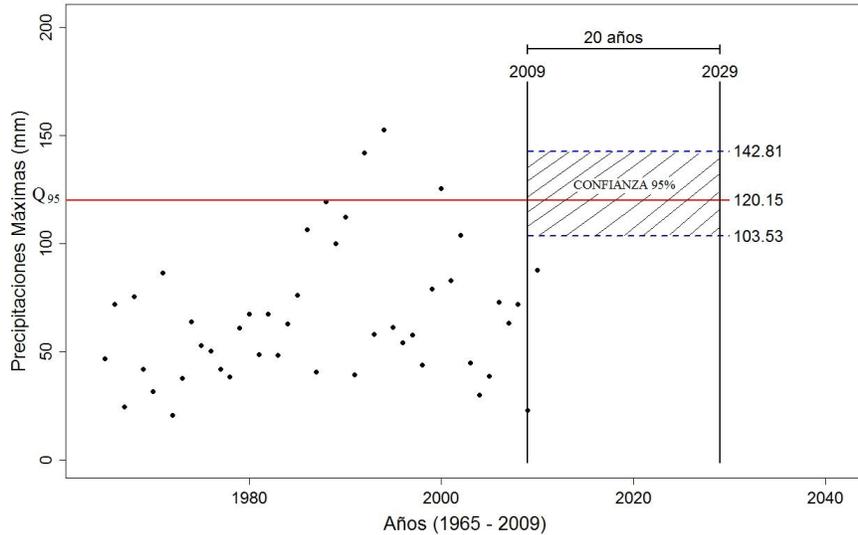


Figura 4.11: Gráfica de confianza para el nivel de retorno $Q_{0.95}$

4.6.2 Vulnerabilidad a inundaciones

El EMV del parámetro de vulnerabilidad a inundaciones θ , considerando $v = 80$ mm, fue $\hat{\theta} = 0.2535$. Es decir, se estima que la probabilidad de que el NPMA sobrepase 80 mm es 0.2535. Nótese que se trata de una probabilidad relativamente alta, indicando una vulnerabilidad significativa de sistemas estructurales asociados a un umbral de $v = 80$ mm. Por otro lado, el intervalo de verosimilitud-confianza del 95% para θ fue $IV(0.15) = [0.1614, 0.3675]$. Es decir, con una confianza del 95%, se estima que la probabilidad de que el NPMA sea mayor que 80 mm se encuentra entre 0.1614 y 0.3675.

Como se mencionó en la Sección 4.2, el parámetro $\theta = P[\text{NPMA} > v]$ obedece a intereses particulares de quienes lo vayan a utilizar, puesto que depende del valor crítico o umbral v del sistema estructural que sea de interés. Por otro lado, en un problema particular se recomienda proporcionar diferentes valores de v , esto con el fin de evaluar la sensibilidad de la estimación de vulnerabilidad ante posibles errores

de especificación de v . Una manera simple de presentar este análisis de sensibilidad se muestra en la Figura 4.12. En esta figura se grafican los niveles de precipitación máximos anuales observados entre los años 1965 y 2009. Además se muestran, para el 2010, las estimaciones puntuales de la vulnerabilidad a inundaciones para diferentes valores de v , dados en la siguiente tabla.

Valor crítico (v)	0	40	60	80	100	120	140
Vulnerabilidad estimada ($\hat{\theta}$)	1	0.809	0.501	0.253	0.116	0.0498	0.0211

Tabla 4.4: EMV del parámetro θ para diferentes valores de v .

Para cada v de la Tabla 4.4, la estimación puntual correspondiente se representa en la Figura 4.12 con un rectángulo vertical ubicado arriba de la estimación, a una altura v . El interior de los rectángulos se pintó en una escala de grises según el correspondiente valor estimado de la vulnerabilidad. En esta escala, el color negro corresponde a $\hat{\theta} = 1$ y el color blanco correspondería a $\hat{\theta} = 0$. Así, se puede trazar una línea imaginaria paralela al eje de las abscisas, a una altura v , con el objetivo de observar de manera conjunta el comportamiento histórico de los niveles de precipitación máximos anuales (con respecto a v) y la probabilidad $\theta = P[\text{NPMA} > v]$ estimada con estos datos para dicho umbral. Por ejemplo, para un valor crítico de $v = 140$ mm, se tienen dos observaciones históricas que sobrepasaron dicho valor y la correspondiente vulnerabilidad estimada fue $\hat{\theta} = 0.0211$, representada con un rectángulo pintado casi de color blanco. Esto contrasta significativamente con la probabilidad estimada $\hat{\theta} = 0.2535$ que se obtiene al considerar el valor $v = 80$ mm, sobrepasado 10 veces a lo largo del periodo bajo estudio.

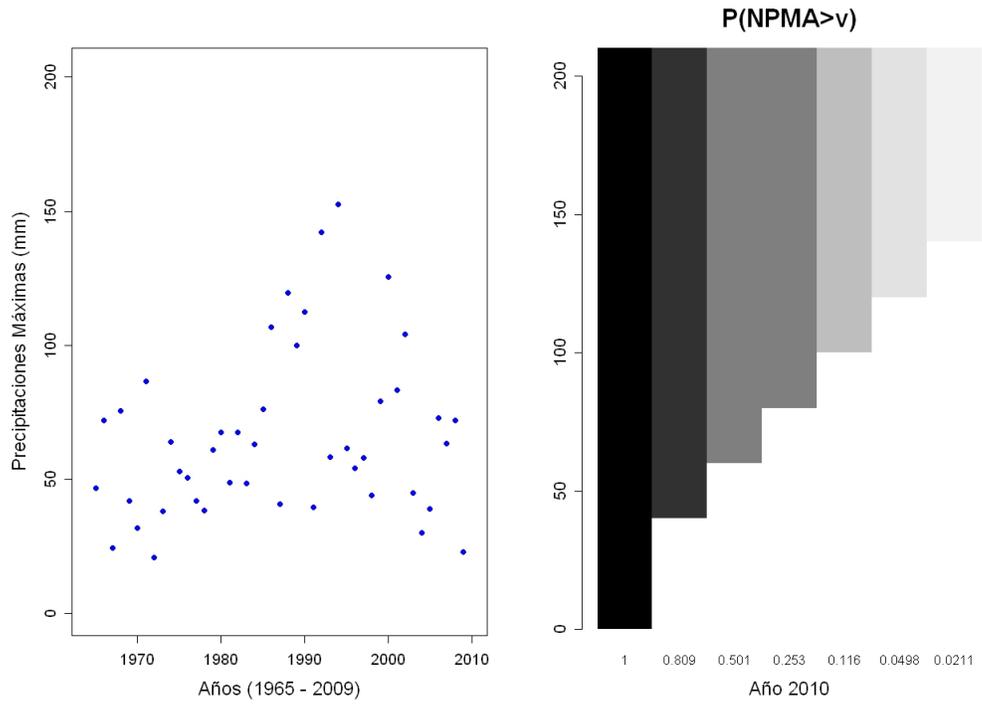


Figura 4.12: Gráfico de probabilidades para el parámetro de vulnerabilidad a inundaciones θ .

Capítulo 5

Conclusiones

Las inundaciones causadas por precipitaciones abundantes son muy comunes y son un problema serio para la sociedad. En este sentido, un problema relevante que se debe abordar de manera científica es el estudio del comportamiento del nivel de precipitación máximo anual (NPMA), que se presentan en forma de lluvia.

En esta tesis se presenta un proceso completo de inferencia estadística (modelación, valoración del modelo, inferencia sobre parámetros de interés en el contexto del problema e interpretación de los resultados en el contexto del problema) para obtener conclusiones relevantes y válidas acerca del comportamiento del NPMA en Hermosillo, Sonora. En particular, se usó la teoría de valores de extremos y el enfoque de verosimilitud para modelar el NPMA en Hermosillo, Sonora, y para hacer inferencia estadística sobre parámetros de interés asociados a la gestión de riesgo contra inundaciones. Se realizaron inferencias sobre un parámetro llamado nivel de retorno y sobre otro parámetro propuesto en esta tesis llamado parámetro de vulnerabilidad a inundaciones.

En relación al caso de estudio analizado en el capítulo anterior se concluye que:

- Dentro de la familia de distribuciones de valores extremos generalizada, el modelo Weibull fue rechazado contundentemente por los datos. En contraste, los datos no proporcionaron evidencia en contra del modelo Gumbel y el modelo Fréchet.

- El modelo Gumbel es un modelo adecuado y parsimonioso para describir el comportamiento probabilístico del NPMA en Hermosillo, Sonora. Por lo tanto, fue utilizado para realizar inferencias sobre el nivel de retorno y el parámetro de vulnerabilidad a inundaciones.
- Se estima que en 20 años se espera observar un NPMA mayor que 120.15 mm.
- Con una confianza del 95%, el NPMA que será sobrepasado una vez en 20 años se encuentra en el intervalo [103.53, 142.81].
- Se estima que en 100 años se espera observar un NPMA mayor que 157.75 mm.
- Con una confianza del 95%, se estima que el NPMA que será sobrepasado una vez en 100 años se encuentra en el intervalo [133.8, 190.57].
- Se estima que la probabilidad de que el NPMA sobrepase 80 mm es 0.2535.
- Con una confianza del 95%, se estima que la probabilidad de que el NPMA sea mayor que 80 mm se encuentra en el intervalo [0.1614, 0.3675].

El proceso completo de inferencia estadística mostrado en esta tesis para el NPMA en Hermosillo, se puede aplicar en otras zonas tanto de Sonora como del país. De hecho, este proceso se puede aplicar a muchos otros problemas relevantes de la sociedad.

Bibliografía

- [1] Agencia Suiza para el desarrollo y la cooperación (2002). Instrumentos de apoyo para el análisis y la gestión de riesgos naturales, *COSUDE-Ayuda Humanitaria y Cuerpo Suizo de Ayuda Humanitaria*, COSUDE. Nicaragua.
- [2] Beirlant, J., Goegebeur, Y. y Teugels, J. (2004). *Statistics of Extremes*. Nueva York: John Wiley & Sons.
- [3] Bolívar, A. M. (2007). *Intervalos de Estimación Bootstrap para Cuantiles en Teoría de Valores Extremos*. Tesis de Maestría en Ciencias con Orientación en Probabilidad y Estadística, CIMAT, Guanajuato, Gto., México.
- [4] Casella, G. y Berger, R. L. (1990). *Statistical Inference*. California: Duxbury Press.
- [5] Chow, V. T., Maidment, D. R. y Mays, I. W. (1994). *Hidrología aplicada*. Colombia: McGraw-Hill.
- [6] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Londres: Springer-Verlag.
- [7] Cox, D. R. y Hinkley, D.V. (1974). *Theoretical Statistics*. Londres: Chapman and Hall.
- [8] Echarri, L. (1998). *Ciencias de la Tierra y del Medio Ambiente*. Madrid: Teide.

- [9] Edwards, A. W. F. (1992). *Likelihood*. Baltimore: The Johns Hopkins University Press.
- [10] Eric, K. y Noji, S. (1997). *The public Health Consequences of Disaster*. Oxford: Oxford University Press.
- [11] Fisher, R. A. (1991). *Statistical Methods and Scientific Inference*. Oxford: Oxford University Press.
- [12] Kalbfleisch, J. G. (1985). *Probability and Statistical Inference*, Vol. 2: Statistical Inference. Nueva York: Springer-Verlag.
- [13] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Nueva York: John Wiley & Sons.
- [14] Lindsey, J. K. (1999). Some Statistical Heresies (With discussion). *The Statistician*, Vol. 48, Págs. 1 - 40.
- [15] Nakamura, M. (2000). *Notas Curso Elementos de Estadística y Probabilidad*, CIMAT, Guanajuato, Gto., México.
- [16] Meeker, W. Q. y Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Nueva York: John Wiley & Sons.
- [17] Montoya, J. A., Díaz-Francés, E. y Sprott, D. A. (2009). On a Criticism of the Profile Likelihood Function. *Statistical Papers*, Vol. 50, Págs. 195-202.
- [18] Montoya, J. A. (2008). *La verosimilitud perfil en la Inferencia Estadística*. Tesis de Doctorado en Ciencias con Orientación en Probabilidad y Estadística, CIMAT, Guanajuato, Gto., México.
- [19] Mood, A. M., Graybill, F. A. y Boes, D. C. (1985). *Introduction to the Theory of Statistics*. Singapore: McGraw-Hill.

- [20] Murray, A. T. y Grubestic, T. H. (2007). *Critical Infrastructure Reliability and Vulnerability*. Nueva York: Springer-Verlag.
- [21] Organización Meteorológica Mundial (2006). Prevención de los desastres y atención a sus defectos, *Informe de un comité de expertos de la OMM 993*, OMM. Suiza.
- [22] Reiss, R. D. y Thomas, M. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Berlin: Birkhäuser Verlag.
- [23] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Nueva York: John Wiley & Sons.
- [24] Sprott, D. A. (2000). *Statistical Inference in Science*. Nueva York: Springer-Verlag.
- [25] Velasco, F. y Hernández, S. (2007). Teoría de valores extremos: Una Introducción. *Revista de ciencias básicas UJAT*, Vol. 6, Págs. 10-16.
- [26] Wanielista, M., Kersten, R. y Eaglin, R. (1997). *Hidrology: water quantity and quality control*. Nueva York: John Wiley & Sons.