



INSTITUTO MEXICANO DEL TRANSPORTE

# Ampliación de funcionalidad de la aplicación *ABAL 0.1* para pruebas de bondad de ajuste

---

Marisol Barrón Bastida  
Eric Moreno Quintero  
Diana Laura Vega Pérez

Publicación Técnica No. 627  
**Sanfandila, Qro.**  
2021

ISSN 0188-7297



Esta investigación fue realizada en la Coordinación de Integración del Transporte del Instituto Mexicano del Transporte, por la Mtra. Marisol Barrón Bastida, el Dr. Eric Moreno Quintero y la Lic. Laura Vega Pérez.

Esta investigación es el producto final del proyecto de investigación interna TI 03/20 Ampliación de funcionalidad de la aplicación ABAL 0.1 para pruebas de bondad de ajuste.

Se agradece la revisión y aportaciones del Dr. Carlos Martner Peyrelongue, coordinador de Transporte Integrado y Logística del IMT, cuyas observaciones mejoraron la calidad de este documento.

Se reconoce también el apoyo brindado por el Laboratorio Nacional de Sistemas de Transporte y Logística (SiT-LOG Lab), cuyas capacidades tecnológicas facilitaron el desarrollo de la aplicación, así como su difusión en el sitio web del SiT-LOG Lab.



# Contenido

---

	Página
Índice de figuras .....	v
Índice de tablas .....	vii
Sinopsis.....	ix
Abstract .....	xi
Resumen ejecutivo.....	xiii
Introducción.....	1
1. Antecedentes .....	3
2. Desarrollo de nuevas funcionalidades de ABAL .....	5
2.1 Módulo de gráficas Q-Q y P-P .....	6
2.2 Módulo distribución Logística.....	7
2.3 Módulo distribución Logarítmica .....	9
2.4 Módulo distribución Gumbel.....	11
2.5 Desarrollo de rutinas de optimización .....	13
2.6 Módulo gráfico para comparar distribuciones .....	15
2.7 Despliegue de nuevas funcionalidades de ABAL.....	16
3. Otras distribuciones de probabilidad .....	21
3.1 Distribución Logarítmica.....	21
3.2 Distribución Logística .....	23
3.3 Distribución Gumbel.....	24
4. La estimación de máxima verosimilitud.....	27
4.1 Distribución Logarítmica.....	28
4.2 Distribución Weibull .....	29
4.3 Distribución Logística .....	31
4.4 Distribución de Gumbel.....	32

---

4.5	Distribución Gamma .....	35
5.	Ejemplos de uso de ABAL con datos reales .....	39
5.1	Ejemplo de la distribución Logarítmica .....	39
5.2	Ejemplo de la distribución Logística .....	41
6.	Resultados y conclusiones.....	47
	Bibliografía .....	51

## Índice de figuras

---

Figura 2.1 Diagrama de caso de uso: modelado de gráficas Q-Q y P-P. ....	7
Figura 2.2 Diagrama de caso de uso de la distribución Logística. ....	9
Figura 2.3 Diagrama de caso de uso de la distribución Logarítmica. ....	11
Figura 2.4 Diagrama de caso de uso de la distribución Gumbel. ....	13
Figura 2.5 Diagrama de caso de uso de la distribución Weibull. ....	14
Figura 2.6 Diagrama de caso de uso de la distribución Gamma. ....	15
Figura 2.7 Diagrama de caso de uso de comparación de distribuciones. ....	16
Figura 2.8 Vista de la distribución Logística, Logarítmica y Gumbel. ....	17
Figura 2.9 Vista de opciones de gráficas Q-Q y P-P. ....	17
Figura 2.10 Vista de las gráficas Q-Q y P-P para una distribución Uniforme. ....	17
Figura 2.11 Vista de las gráficas Q-Q y P-P para una distribución Normal. ....	18
Figura 2.12 Vista de las gráficas Q-Q y P-P para una distribución Exponencial. ...	18
Figura 2.13 Vista de las gráficas Q-Q y P-P para una distribución Lognormal. ....	18
Figura 2.14 Vista de las gráficas Q-Q y P-P para una distribución Weibull. ....	19
Figura 2.15 Vista de las gráficas Q-Q y P-P para una distribución Gamma. ....	19
Figura 2.16 Vista de las gráficas de comparación de distribuciones. ....	20
Figura 3.1 Función de probabilidad Logarítmica. ....	22
Figura 3.2 Densidad de probabilidad Logística. ....	24
Figura 3.3 Densidad de probabilidad Gumbel. ....	25
Figura 4.1 Aproximación gráfica al parámetro $\theta$ . ....	28
Figura 4.2 Estimación del parámetro $\theta$ con Solver de Excel. ....	29
Figura 4.3 Estimación de parámetros Weibull $\lambda, k$ con Solver de Excel. ....	30

Figura 4.4 Solución MLE de parámetros Weibull $\lambda, k$ con Solver de Excel. ....	30
Figura 4.5 Estimación de parámetros $a, b$ de la distribución Logística.....	31
Figura 4.6 Parámetros de máxima verosimilitud $a, b$ en distribución Logística. ....	32
Figura 4.7 Aproximación gráfica al parámetro $b$ Gumbel. ....	34
Figura 4.8 Estimación de parámetros $a, b$ de la distribución Gumbel. ....	34
Figura 4.9 Parámetros de máxima verosimilitud $a, b$ en distribución Gumbel. ....	35
Figura 4.10 Estimación de parámetros $\alpha, \beta$ de la distribución Gamma. ....	36
Figura 4.11 Estimación de parámetros $\alpha, \beta$ de la distribución Gamma. ....	37
Figura 4.12 Solución de Solver. ....	37
Figura 5.1 Prueba Ji-cuadrada para distrib. Logarítmica.....	40
Figura 5.2 Prueba Histograma de datos y ajuste de Logarítmica $\theta = 0.55263$ .....	41
Figura 5.3 Estimación de parámetros $a, b$ de la Logística.....	42
Figura 5.4 Gráfica Q-Q de datos de exportación de algodón ajustados a una distribución Logística.....	43
Figura 5.5 Estimación de parámetros $a, b$ de la Gumbel. ....	44
Figura 5.6 Gráfica Q-Q de datos de daños materiales en accidentes carreteros ajustados a una distribución Gumbel. ....	45
Figura 6.1 Enfatizar la categoría del eje.....	48
Figura 6.2 Diagrama de pasos. ....	48



## Índice de tablas

---

Tabla 1.1 Distribuciones y tests.....	4
Tabla 2.1 Requerimientos de interfaz de gráficos Q-Q y P-P.....	6
Tabla 2.2 Requerimientos funcionales de gráficos Q-Q y P-P. ....	6
Tabla 2.3 Requerimientos de interfaz de la distribución Logística.....	7
Tabla 2.4 Requerimientos funcionales de la distribución Logística. ....	8
Tabla 2.5 Requerimiento de interfaz de la distribución Logarítmica.....	9
Tabla 2.6 Requerimientos funcionales de la distribución Logarítmica.....	10
Tabla 2.7 Requerimientos de interfaz de la distribución Gumbel. ....	11
Tabla 2.8 Requerimientos funcionales de la distribución Gumbel.....	12
Tabla 2.9 Requerimientos funcionales de la distribución Weibull.....	13
Tabla 2.10 Requerimientos funcionales de la distribución Gamma.....	14
Tabla 2.11 Requerimientos de interfaz de gráficos para la comparación de distribuciones. ....	15
Tabla 5.1 Núm. de tripulantes de vehículos utilitarios. Estación 246 “Playas de Tijuana” EECAN 2010.....	39
Tabla 5.2 Exportación de algodón a EUA por Laredo, 2019 (tons). ....	41
Tabla 5.3 Daños materiales (miles de dólares) por accidentes en carreteras federales por estado en 2018.....	43



## Sinopsis

---

Este trabajo presenta las nuevas funcionalidades de ABAL (Aplicación para Bondad de Ajuste en Línea). Estas incluyen nuevas distribuciones de probabilidad y la utilización del método de máxima verosimilitud (MLE) para la estimación de parámetros, para reemplazar los casos tratados con el método de momentos en la primera versión de ABAL. La estimación MLE utiliza procedimientos de optimización no-lineal los cuales son explicados en este trabajo. Además, se incorporaron funciones gráficas como: gráficas Q-Q, P-P y comparación de distribuciones.

Con estas nuevas funciones, el usuario tendrá una variedad de funciones para visualizar y analizar sus pruebas estadísticas, y con ello decidir si su muestra de datos puede representarse adecuadamente por alguna ley de probabilidad elegida. Se espera que los comentarios recibidos por los usuarios enriquezcan y mejoren la funcionalidad de ABAL.

Para el desarrollo de las nuevas funcionalidades se usaron las librerías: *Seaborn*, *Statsmodels*, *Bokeh* y *Matplotlib*. Dichas funciones se adicionaron a la sección de ayuda de ABAL, actualmente disponible en la dirección: <http://abal.imt.mx/>



# Abstract

---

This work presents the new functions of ABAL (Application for Goodness of Fit Online). These include new probability distributions and maximum likelihood estimation of parameters in order to replace the cases using the method of moments in the first version of ABAL. MLE estimation uses non-linear procedures, explained in this work. In addition, graphical functions such as: Q-Q graphs, P-P graphs, and comparison of distributions were incorporated.

With these new features, the user will have a variety of functions to visualize and analyze his(her) statistical tests, and thereby decide if his(her) data sample can be adequately represented by some chosen probability law. Feedback from users is expected to enrich and improve ABAL's functionality.

For the development of the new functionalities, the libraries: Seaborn, Statsmodels, Bokeh and Matplotlib were used. ABAL is available at the address: <http://abal.imt.mx/>.



# Resumen ejecutivo

---

Este trabajo presenta el desarrollo de nuevas funcionalidades de ABAL, aplicación informática en línea que realiza pruebas de bondad de ajuste que se complementan con una variedad de gráficos para la visualización y análisis de los datos de entrada como: gráficos Q-Q, P-P y comparación de distribuciones.

Las pruebas de bondad de ajuste de ABAL se desarrollaron con el método de máxima verosimilitud para estimar los parámetros, tanto para las nuevas funciones de probabilidad agregadas como para las que usaban el método de momentos en la primera versión de ABAL. El método de máxima verosimilitud es más preciso y estable numéricamente que el método de momentos y permite decidir mejor si un modelo probabilístico es adecuado dadas las observaciones del usuario.

En el capítulo 1 “Antecedentes” se mencionan los trabajos previos realizados en el desarrollo de la aplicación ABAL, como una etapa subsecuente de la Publicación Técnica No. 545 “Modelos de probabilidad en transporte e ingeniería. Usos comunes y ajuste de datos” y la Publicación Técnica No. 577 “Aplicación Informática para pruebas estadísticas de bondad de ajuste”. Se describen las distribuciones y pruebas de ajuste disponibles, así como los métodos utilizados para su desarrollo.

El capítulo 2, “Desarrollo de nuevas funcionalidades de ABAL”, describe el desarrollo de las nuevas características y la definición de sus requerimientos integrados en la aplicación. Estos se describen a través del modelado de diagramas UML (Lenguaje Unificado de Modelado). Así mismo, se muestra el prototipo de diseño de las gráficas incorporadas.

En el capítulo 3, “Otras distribuciones de probabilidad”, se describen las tres distribuciones que fueron agregadas a ABAL, estas son: Logarítmica, Logística y Gumbel. Esta descripción incluye: el cálculo de los parámetros a utilizar, su función de probabilidad, su uso en trabajos de investigación y una gráfica como ejemplo.

En el capítulo 4 “La estimación de máxima verosimilitud”, se muestra el desarrollo de estimadores de máxima verosimilitud para las distribuciones de ABAL y los planteamientos de programación no-lineal requeridos en cada caso, así como su solución. Las distribuciones que se incluyen son Logarítmica, Weibull, Logística, Gumbel y Gamma.

En el capítulo 5, “Ejemplos de uso de ABAL con datos reales”, se muestran algunos casos con datos reales de distintas fuentes para ejemplificar nuevas distribuciones que se agregaron a ABAL, como la distribución Logarítmica del tipo discreto y las distribuciones continuas Logística y Gumbel.

El capítulo 6 “Resultados y conclusiones” resume el desarrollo de las nuevas funcionalidades de ABAL, partiendo de la necesidad de mejorar la estimación de los parámetros de las distribuciones con el uso de estimadores de máxima verosimilitud, indicando además las nuevas gráficas que permite realizar ABAL. Así mismo, se identificaron algunas líneas de trabajo a futuro.







# Introducción

---

Este trabajo presenta el desarrollo de las nuevas funcionalidades de ABAL una aplicación informática en línea que realiza pruebas de bondad de ajuste para los datos de entrada de los usuarios. Estas funcionalidades consisten en la integración de tres distribuciones más: la distribución Logística, la Logarítmica y la Gumbel; el uso de máxima verosimilitud para la estimación de parámetros y nuevas gráficas de utilidad para evaluar la calidad del ajuste de los datos.

Las rutinas de cálculo utilizan el método de máxima verosimilitud debido a que es mucho más estable y da resultados de mayor precisión. Por este motivo, en la segunda etapa de la aplicación ABAL se ha buscado reemplazar el método de momentos en las distribuciones que lo permitan, y extenderlo a las nuevas distribuciones de probabilidad que se incluyen en esta segunda etapa de ABAL.

La estimación de parámetros por el método de momentos tiene la ventaja de tener un manejo algebraico sencillo, sin embargo, al tratar con muestras pequeñas, se dan casos en que el método de momentos puede dar resultados fuera del rango esperado para los parámetros que estima.

El método de máxima verosimilitud, por otra parte, es mucho más estable y da resultados de mayor precisión, pero tiene la desventaja que el tratamiento analítico puede llevar a sistemas de ecuaciones simultáneas no-lineales muy difíciles de resolver o sencillamente intratables algebraicamente, exigiendo métodos numéricos y de optimización adecuados para su solución.

En la parte de visualización, entre las nuevas funcionalidades se añadió la capacidad de generar gráficas Q-Q (cuantil-cuantil) y gráficas P-P (probabilidad-probabilidad) que permiten rápidamente una evaluación gráfica de la bondad del ajuste de los datos. Las primeras permiten visualizar las comparaciones entre los cuantiles calculados de la muestra de datos, con los cuantiles teóricos del modelo de probabilidad probado, mientras que las segundas comparan los valores de probabilidad empírica a partir de la muestra con los de la probabilidad teórica del modelo probado.

En el desarrollo de gráficas, también se incluye un módulo para graficar distribuciones de probabilidad variando parámetros o comparando modelos, para tener una visualización que ayude al usuario a contrastar sus hipótesis de trabajo.

En los capítulos siguientes se muestran estos desarrollos.



# 1. Antecedentes

---

En el primer trimestre de 2019 el IMT liberó la Publicación Técnica No. 545 “Modelos de probabilidad en transporte e ingeniería. Usos comunes y ajuste de datos”, en la que se documentan y resumen las técnicas estadísticas de ajuste de datos que se requieren para tener un modelo de probabilidad adecuado para los datos de interés.

Así mismo, a finales del primer semestre del 2020 la Coordinación de Transporte Integrado y Logística del IMT liberó la Publicación Técnica No. 577 “Aplicación Informática para pruebas estadísticas de bondad de ajuste”, la cual detalla el desarrollo de la aplicación informática ABAL (Aplicación para Bondad de Ajuste en Línea).

En el contenido de dicha publicación se menciona que como resultado se obtuvo la versión 0.1 de ABAL, que básicamente fue el primer lanzamiento en la Web de prueba con los usuarios. El título de la presente publicación hace referencia a dicha versión y con la ampliación del desarrollo de ABAL se actualizará su versión a la 1.0.

La aplicación ABAL se encuentra disponible en línea en la dirección <http://abal.imt.mx> y es de libre acceso para cualquier usuario. Resuelve de manera práctica y eficiente la realización de pruebas de bondad de ajuste; además, genera reportes en formato PDF con los resultados del cálculo y sus complementos gráficos (histogramas).

En la primera versión, se pueden encontrar las distribuciones de probabilidad de uso más común en proyectos de investigación y de solución de problemas en ingeniería y transporte. Estas distribuciones son: Uniforme, Normal, Exponencial, Lognormal, Weibull, Gamma, Binomial, Geométrica (tanto para núm. de fallas como para núm. de ensayos) y Poisson.

La estimación de parámetros en la primera versión ABAL utiliza el método de momentos en los casos de: la Uniforme, la Gamma y la Binomial.

Con la distribución Weibull el método de momentos se dificulta; así que, con una transformación logarítmica de la función de distribución acumulada, se obtiene una forma lineal que admite una regresión lineal para estimar parámetros.

En las demás distribuciones se usa el método de máxima verosimilitud para la estimación de los parámetros.

La tabla 1.1 muestra los tests que ofrece ABAL para cada una de las distribuciones.

**Tabla 1.1 Distribuciones y tests**

	Ji-cuadrada	Kolmogorov-Smirnov	Anderson-Darling
Uniforme	✓	✓	
Normal	✓	✓	✓
Exponencial	✓	✓	✓
Lognormal	✓	✓	
Weibull	✓	✓	
Gamma	✓	✓	
Binomial	✓		
Geométrica Núm. de fallas	✓		
Geométrica Núm. de ensayos	✓		
Poisson	✓		

Fuente: Elaboración propia.

En el desarrollo de ABAL, la estimación de parámetros a partir de la muestra del usuario, en algunos casos se usó el método de momentos, que iguala los momentos teóricos del modelo probabilístico con los momentos muestrales, y en otros casos, donde había disponible una fórmula directa, se usó el método de máxima verosimilitud.

## 2. Desarrollo de nuevas funcionalidades de ABAL

---

El presente capítulo integra los módulos de desarrollo de las nuevas funcionalidades de ABAL. Dichos módulos definen los requerimientos que se integrarán a la aplicación, estos son:

- Desarrollo de gráficas Q-Q y P-P.
- Desarrollo de las distribuciones: Logística, Gumbel y Logarítmica.
- Comparación de la distribución en gráficos.
- Desarrollo de rutinas de optimización.
- Subir en línea ejemplos de muestras resueltas con ABAL.
- Actualizar el módulo de ayuda.

De acuerdo al Lenguaje Unificado de Modelado (UML), se exponen las nuevas funcionalidades de ABAL. En cada una de estas se listan sus requisitos correspondientes al diseño de interfaz y funcionalidad interna. Así mismo, cada funcionalidad es modelada en casos de uso como un conjunto de entradas, comportamientos y salidas. Finalmente, se ilustran las funcionalidades en prototipos de interfaces.

Es importante citar que, la metodología de gestión de requerimientos de Google, menciona que, “para sistemas grandes y complejos es imposible cumplir con los requerimientos de consistencia y completión. La razón de esto se debe parcialmente a la complejidad inherente del sistema y parcialmente a que los diferentes puntos de vista de los usuarios tienen necesidades inconsistentes. Los problemas emergen después de un análisis profundo”.

En este sentido, con la especificación de los nuevos requerimientos se espera asegurar la completión<sup>1</sup> y consistencia<sup>2</sup> de ABAL, sin embargo, una vez que los

---

<sup>1</sup> Significa que todos los servicios solicitados por el usuario están definidos.

<sup>2</sup> Significa que todos los requerimientos no tienen funciones contradictorias.

requerimientos sean descubiertos en las diferentes revisiones o en las fases posteriores del ciclo de vida de ABAL, se procederá a su actualización.

## 2.1 Módulo de gráficas Q-Q y P-P

En este apartado se presentan las especificaciones del caso de uso del módulo de gráficas Q-Q y P-P. El caso describe los datos de entrada y salida del requerimiento. Cabe señalar que este proceso de cálculo se integrará únicamente a las distribuciones continuas. En la tabla 2.1 y 2.2 se describen los requisitos necesarios para cumplir con este módulo.

**Tabla 2.1 Requerimientos de interfaz de gráficos Q-Q y P-P.**

<b>Número de requisito</b>	RI-1		
<b>Nombre de requisito</b>	Opción de cálculo de gráficos Q-Q y P-P		
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción	
<b>Fuente del requisito</b>	Equipo de trabajo		
<b>Descripción del requisito</b>	Agregar la opción de cálculo de gráficos Q-Q y P-P		
Agregar en la sección “genera histograma”: <ul style="list-style-type: none"> <li>• Un cuadro de opción (check-box) como opción de selección de cálculo: gráficos Q-Q.</li> <li>• Un cuadro de opción (check-box) como opción de selección de cálculo: gráficos P-P.</li> <li>• Un área de vista para ambos gráficos.</li> </ul>			
<b>Prioridad del requisito</b>	<input checked="" type="checkbox"/> Alta/Eencial	<input type="checkbox"/> Media/Deseado	<input type="checkbox"/> Baja/ Opcional

Fuente: Elaboración propia.

**Tabla 2.2 Requerimientos funcionales de gráficos Q-Q y P-P.**

<b>Número de requisito</b>	RF- 1		
<b>Nombre de requisito</b>	Desarrollo de gráficos Q-Q y P-P.		
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción	
<b>Fuente del requisito</b>	Equipo de trabajo		
<b>Descripción del requisito</b>	Desarrollar los métodos de cálculo para gráficos Q-Q y P-P en todas las distribuciones hasta ahora disponibles en ABAL.		
<ul style="list-style-type: none"> <li>• Como datos de entrada se requiere únicamente la muestra.</li> <li>• Desarrollar el proceso de cálculo del Q-muestral (probabilidad correspondiente a la distribución)</li> <li>• Obtener los parámetros correspondientes a la distribución y con ello, desarrollar el proceso de cálculo de la Q ~Distribución.</li> </ul>			
<b>Prioridad del requisito</b>	<input checked="" type="checkbox"/> Alta/Eencial	<input type="checkbox"/> Media/Deseado	<input type="checkbox"/> Baja/ Opcional

Fuente: Elaboración propia.

La figura 2.1 ilustra el modelado del caso de uso de gráficos Q-Q y P-P. En este se observa el flujo de eventos entre el usuario (actor) y la aplicación web, así como las acciones o actividades que lo conforman.



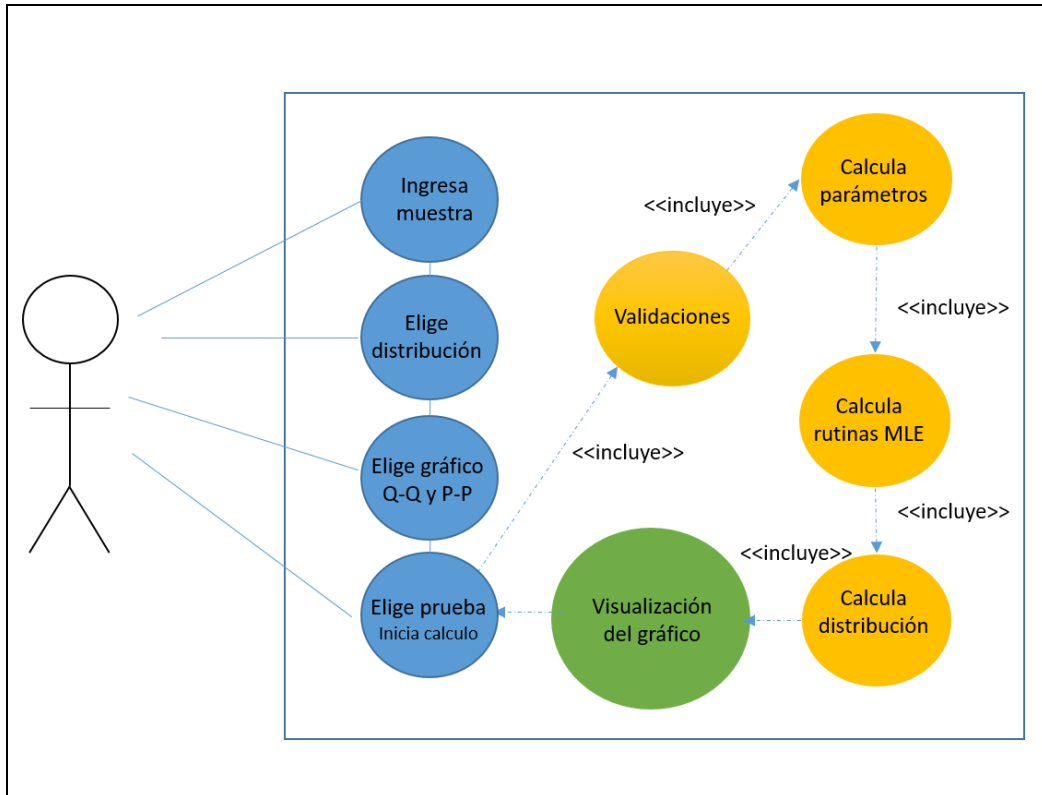


Figura 2.1 Diagrama de caso de uso: modelado de gráficas Q-Q y P-P.

## 2.2 Módulo distribución Logística

En este apartado se presentan las especificaciones del caso de uso del módulo de la distribución Logística, dicho caso incluye la descripción de los datos de entrada y salida del requerimiento.

En las tablas 2.3 y 2.4 se definen los requerimientos de la interfaz y el cálculo de las rutinas de optimización respectivamente. En la figura 2.2 se observa el caso de uso que muestra el flujo de eventos de los procesos de cálculo de la distribución Logística.

Tabla 2.3 Requerimientos de interfaz de la distribución Logística.

<b>Número de requisito</b>	RI-2	
<b>Nombre de requisito</b>	Requisitos de interfaz para la distribución Logística	
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo	
<b>Descripción del requisito</b>	Actualización de la página de inicio de ABAL y agregar una nueva página para el cálculo de la distribución Logística.	
<ol style="list-style-type: none"> <li>1. En el apartado de distribuciones de probabilidad continuas, agregar la distribución Logística</li> <li>2. Agregar una nueva página para el cálculo de la distribución Logística, con las siguientes secciones: <ul style="list-style-type: none"> <li>• Las fórmulas que definan la distribución</li> </ul> </li> </ol>		

<ul style="list-style-type: none"> <li>• Recomendación del número de datos mínimos a ingresar para obtener una prueba confiable.</li> <li>• Tabla para ingresar la muestra</li> <li>• Sección de “genera histograma”</li> <li>• Opción de comparación de la distribución en gráficos</li> <li>• Opción de cálculo de gráficos Q-Q y P-P</li> <li>• Opción de cálculo de tests: Ji-Cuadrada y Kolmogorov-Smirnov</li> <li>• Sección de resultados con: gráficos, datos y descarga de resultados en PDF.</li> </ul>
<b>Prioridad del requisito</b> <input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaboración propia.

**Tabla 2.4 Requerimientos funcionales de la distribución Logística.**

<b>Número de requisito</b>	RF-2	
<b>Nombre de requisito</b>	Requisitos funcionales para la distribución Logística	
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo	
<b>Descripción del requisito</b>	Desarrollar los métodos y clases para el cálculo de la distribución Logística a partir de Estimadores de Máxima Verosimilitud.	
	<ol style="list-style-type: none"> <li>1. A partir de los parámetros <math>a</math> y <math>b</math> desarrollar una rutina para el cálculo de la función objetivo a maximizar</li> <li>2. A partir de una función de optimización, desarrollar una rutina para maximizar la función objetivo y con ello obtener los estimadores MLE necesarios para esta distribución.             <ul style="list-style-type: none"> <li>• En dicha rutina restringir el valor de <math>b</math> como <math>&gt;0.0001</math> para asegurar que el algoritmo de optimización no llegue a usar el valor de cero en sus rutinas de cálculo.</li> <li>• Una vez obtenidos los valores de los parámetros <math>a</math> y <math>b</math> desarrollar un proceso para el cálculo de la prueba Ji-cuadrada.</li> </ul> </li> </ol>	
<b>Prioridad del requisito</b>	<input checked="" type="checkbox"/> Alta/Esencial	<input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaboración propia.

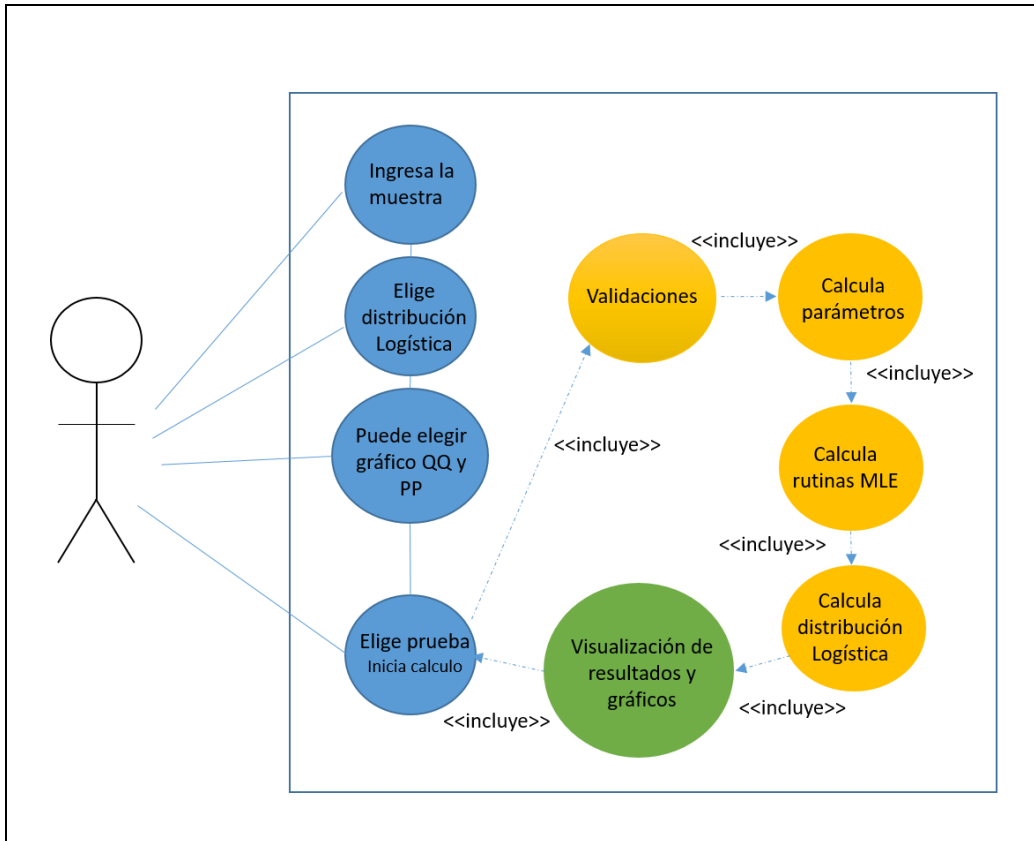


Figura 2.2 Diagrama de caso de uso de la distribución Logística.

### 2.3 Módulo distribución Logarítmica

En este apartado se presentan las especificaciones del caso de uso del módulo de la distribución Logarítmica, dicho caso incluye la descripción de los datos de entrada y salida del requerimiento.

En las tablas 2.5 y 2.6 se definen los requerimientos de la interfaz y el cálculo de las rutinas de optimización respectivamente. En la figura 2.3 se observa el caso de uso que muestra el flujo de eventos de los procesos de cálculo de la distribución Logarítmica.

Tabla 2.5 Requerimiento de interfaz de la distribución Logarítmica.

<b>Número de requisito</b>	RI-3	
<b>Nombre de requisito</b>	Requisitos de interfaz para la distribución Logarítmica	
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo	
<b>Descripción del requisito</b>	Actualización de la página de inicio de ABAL y agregar una nueva página para el cálculo de la distribución Logarítmica.	
1. En el apartado de distribuciones de probabilidad discreta, agregar la distribución Logarítmica. 2. Agregar una nueva página para el cálculo de la distribución Logarítmica con las siguientes secciones: <ul style="list-style-type: none"> <li>Las fórmulas que definan la distribución</li> </ul>		

<ul style="list-style-type: none"> <li>• Recomendación del número de datos mínimos a ingresar para obtener una prueba confiable.</li> <li>• Tabla para ingresar la muestra</li> <li>• Sección de “genera histograma”</li> <li>• Opción de comparación de la distribución en gráficos</li> <li>• Opción de cálculo de gráficos Q-Q y P-P</li> <li>• Opción de cálculo de tests: Ji-Cuadrada y Kolmogorov-Smirnov</li> <li>• Sección de resultados con: gráficos, datos y descarga de resultados en PDF.</li> </ul>
<b>Prioridad del requisito</b> <input checked="" type="checkbox"/> Alta/Eencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaboración propia.

**Tabla 2.6 Requerimientos funcionales de la distribución Logarítmica.**

<b>Número de requisito</b>	RF-3	
<b>Nombre de requisito</b>	Requisitos funcionales para la distribución Logarítmica	
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo	
<b>Descripción del requisito</b>	Desarrollar los métodos y clases para el cálculo de la distribución Logarítmica a partir de Estimadores de Máxima Verosimilitud.	
	<ol style="list-style-type: none"> <li>1. A partir del valor inicial del estimador <math>\hat{\theta}</math> desarrollar una rutina para el cálculo de la función objetivo a minimizar.</li> <li>2. A partir de una función de optimización, desarrollar una rutina para minimizar la función objetivo, y con ello cambiar el valor inicial de <math>\hat{\theta}</math> . De esta forma se logra obtener la función objetivo igual a cero o lo más cercano posible.</li> <li>3. Una vez obtenido el valor de <math>\hat{\theta}</math> desarrollar una rutina para el cálculo de la prueba Ji-cuadrada.</li> </ol>	
<b>Prioridad del requisito</b>	<input checked="" type="checkbox"/> Alta/Eencial	<input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaboración propia.

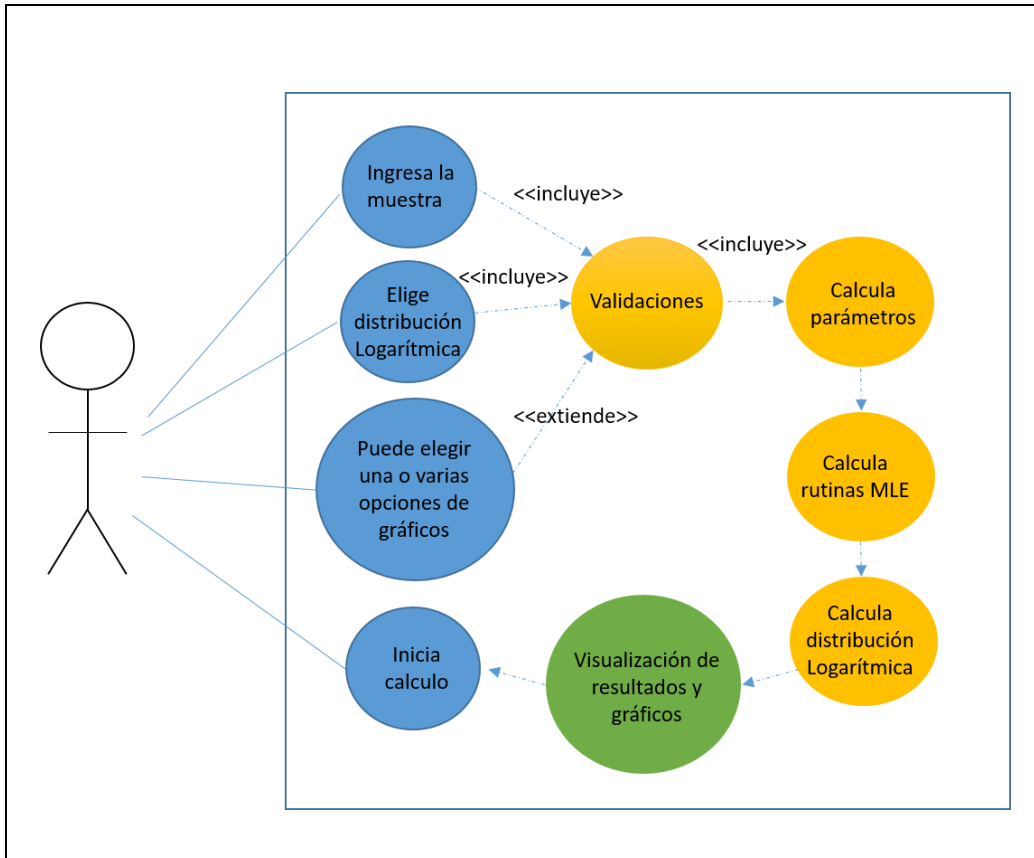


Figura 2.3 Diagrama de caso de uso de la distribución Logarítmica.

## 2.4 Módulo distribución Gumbel

En este apartado se presentan las especificaciones del caso de uso del módulo de la distribución Gumbel, dicho caso incluye la descripción de los datos de entrada y salida del requerimiento.

En las tablas 2.7 y 2.8 se definen los requerimientos de la interfaz y el cálculo de las rutinas de optimización respectivamente. En la figura 2.4 se observa el caso de uso que muestra el flujo de eventos de los procesos de cálculo de la distribución Gumbel.

Tabla 2.7 Requerimientos de interfaz de la distribución Gumbel.

<b>Número de requisito</b>	RI-4
<b>Nombre de requisito</b>	Requisitos de interfaz para la distribución Gumbel.
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito <input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo
<b>Descripción del requisito</b>	Actualización de la página de inicio de ABAL y agregar una nueva página para el cálculo de la distribución Gumbel.
1. En el apartado de distribuciones de probabilidad continua agregar la distribución Gumbel. 2. Agregar una nueva página para el cálculo de la distribución Gumbel con las siguientes secciones:	

<ul style="list-style-type: none"> <li>• Las fórmulas que definan la distribución</li> <li>• Recomendación del número de datos mínimos a ingresar para obtener una prueba confiable.</li> <li>• Tabla para ingresar la muestra</li> <li>• Sección “genera histograma”</li> <li>• Opción de comparación de la distribución en gráficos</li> <li>• Opción de cálculo de gráficos Q-Q y P-P</li> <li>• Opción de cálculo de tests: Ji-Cuadrada y Kolmogorov-Smirnov</li> <li>• Sección de resultados con: gráficos, datos y descarga de resultados en PDF.</li> </ul>
<b>Prioridad del requisito</b> <input checked="" type="checkbox"/> Alta/Esencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaborado por el IMT

**Tabla 2.8 Requerimientos funcionales de la distribución Gumbel.**

<b>Número de requisito</b>	RF- 4	
<b>Nombre de requisito</b>	Requisitos funcionales para la distribución Gumbel.	
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo	
<b>Descripción del requisito</b>	Desarrollar los métodos y clases para el cálculo de la distribución Gumbel a partir de Estimadores de Máxima Verosimilitud.	
	<ol style="list-style-type: none"> <li>1. Desarrollar una rutina para estimar el valor inicial del parámetro <math>b</math>.</li> <li>2. Desarrollar una rutina de optimización para minimizar la función objetivo, y con ello cambiar el valor inicial del parámetro <math>b</math>. De esta forma se logra obtener la función objetivo igual a cero o lo más cercano posible.</li> <li>3. Una vez obtenido el valor de <math>b</math> desarrollar una rutina para el cálculo de la prueba Ji-cuadrada.</li> </ol>	
<b>Prioridad del requisito</b>	<input checked="" type="checkbox"/> Alta/Esencial	<input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaboración propia.

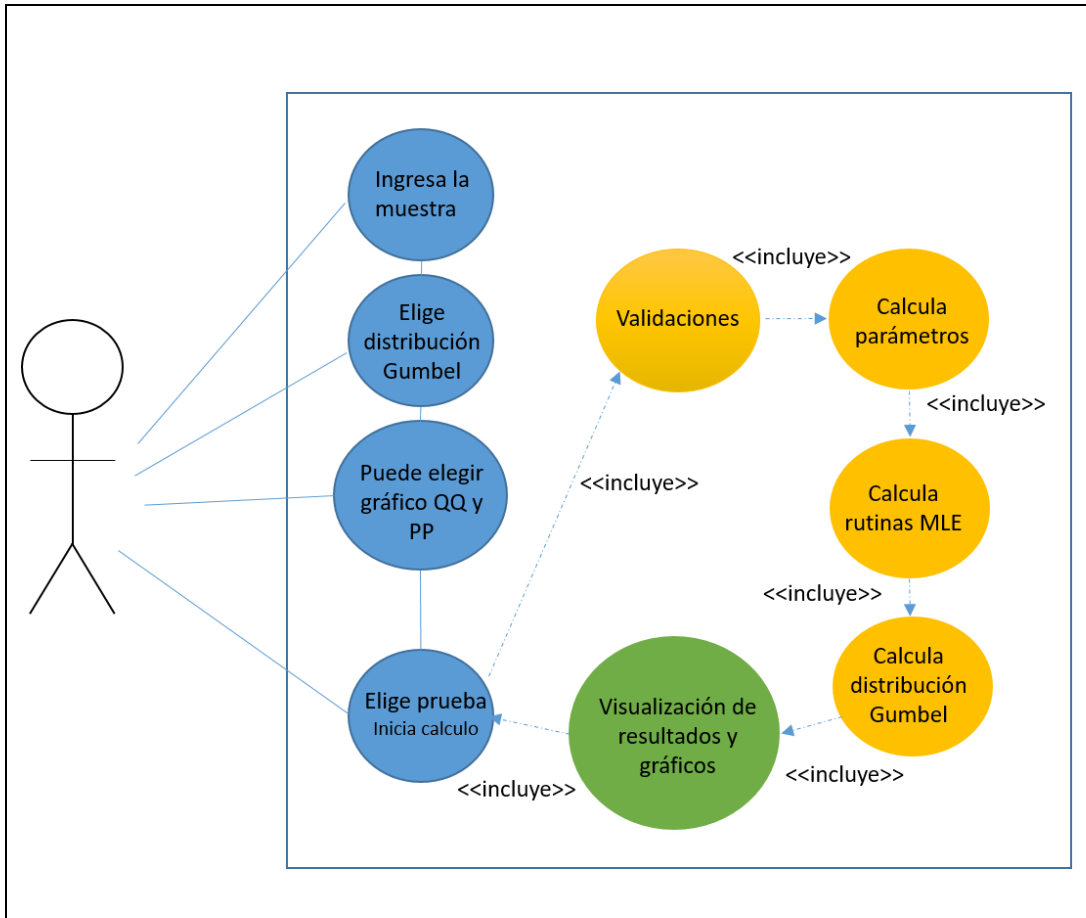


Figura 2.4 Diagrama de caso de uso de la distribución Gumbel.

## 2.5 Desarrollo de rutinas de optimización

En este apartado se presentan las especificaciones de los casos de uso para el desarrollo de rutinas de optimización. El objetivo de desarrollar estas rutinas consiste en encontrar los estimadores MLE para el cálculo del test de dichas distribuciones. De esta manera se consigue actualizar los cálculos y con ello obtener un resultado óptimo.

En la tabla 2.9 se definen los requisitos para el cálculo de rutinas de optimización. En la imagen 2.5 se observa el caso de uso que muestra el flujo de eventos de los procesos de cálculo para la distribución Weibull.

Tabla 2.9 Requerimientos funcionales de la distribución Weibull.

<b>Número de requisito</b>	RF-5	
<b>Nombre de requisito</b>	Requisitos funcionales para la distribución Weibull	
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo	
<b>Descripción del requisito</b>	Desarrollar los métodos y clases para el cálculo de rutinas de optimización	

1. A partir de los valores iniciales de  $\lambda_0, k_0$  desarrollar una rutina para el cálculo de la función objetivo.
2. Desarrollar una rutina de optimización para maximizar la función objetivo, y con ello encontrar los estimadores MLE para calcular la distribución.
  - Obtener la función objetivo igual a cero o lo más cercano posible.

**Prioridad del requisito**     Alta/Eencial     Media/Deseado     Baja/ Opcional

Fuente:    Elaboración propia.

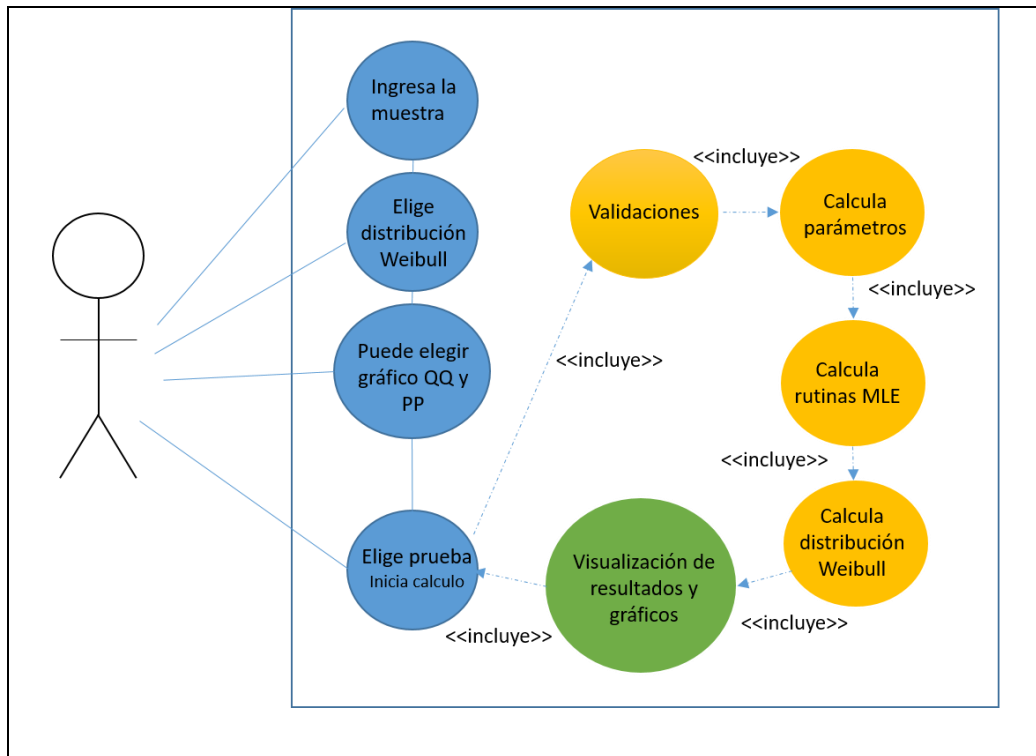


Figura 2.5 Diagrama de caso de uso de la distribución Weibull.

En la tabla 2.10 se definen los requisitos para el cálculo de rutinas de optimización. En la figura 2.6 se observa el caso de uso que muestra el flujo de eventos de los procesos de cálculo para la distribución Gamma.

Tabla 2.10 Requerimientos funcionales de la distribución Gamma.

<b>Número de requisito</b>	RF-6	
<b>Nombre de requisito</b>	Requisitos funcionales para la distribución Gamma.	
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito	<input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo	
<b>Descripción del requisito</b>	Desarrollar los métodos y clases para el cálculo de rutinas de optimización	
<ol style="list-style-type: none"> <li>1. A partir de los valores iniciales de <math>\alpha, \beta</math> desarrollar una rutina para el cálculo de la función objetivo.</li> <li>2. Desarrollar una rutina de optimización para maximizar la función objetivo y con ello encontrar los estimadores MLE para calcular la distribución.                     <ul style="list-style-type: none"> <li>• Obtener la función objetivo igual a cero o lo más cercano posible.</li> </ul> </li> </ol>		



<b>Prioridad del requisito</b> <input checked="" type="checkbox"/> Alta/Eencial <input type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaboración propia.

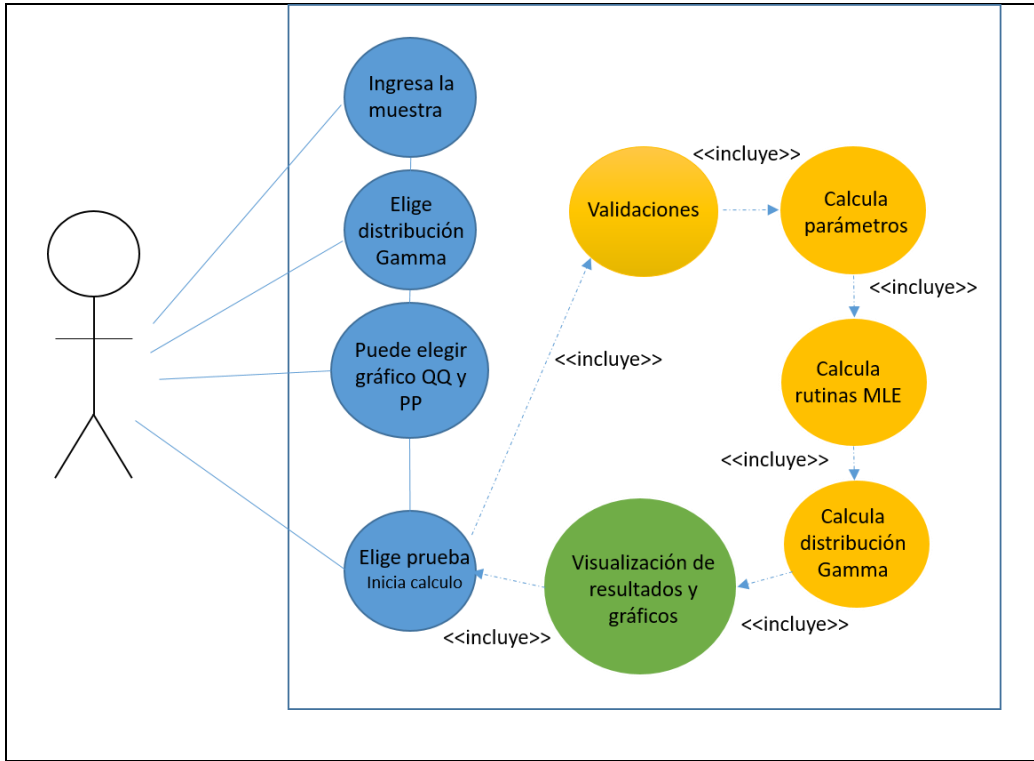


Figura 2.6 Diagrama de caso de uso de la distribución Gamma.

## 2.6 Módulo gráfico para comparar distribuciones

En este apartado se presentan las especificaciones del caso de uso para el proceso de comparación de distribuciones, la cuales, serán visualizadas en diversos gráficos. Las opciones de gráficos disponibles son: distribución individual, familia paramétrica, área bajo la curva y dos distribuciones. En la tabla 2.11 se definen los requisitos para el cálculo de los gráficos y en la figura 2.7 se describe el flujo de eventos del proceso de cálculo.

Tabla 2.11 Requerimientos de interfaz de gráficos para la comparación de distribuciones.

<b>Número de requisito</b>	RI-5
<b>Nombre de requisito</b>	Comparación de distribuciones
<b>Tipo</b>	<input checked="" type="checkbox"/> Requisito <input type="checkbox"/> Restricción
<b>Fuente del requisito</b>	Equipo de trabajo
<b>Descripción del requisito</b>	Agregar nuevas opciones de gráficos para la comparación de distribuciones.

1. Agregar una sección de opciones de gráficos que incluya:

<ul style="list-style-type: none"> <li>• Distribución individual</li> <li>• Familia paramétrica</li> <li>• Área bajo la curva</li> <li>• Dos distribuciones</li> </ul> <ol style="list-style-type: none"> <li>2. Para generar el gráfico de “distribución individual” se requieren: la distribución y los parámetros que correspondan a la distribución.</li> <li>3. Para el gráfico de “familia paramétrica” se requiere: la distribución y el conjunto de parámetros. Se podrán visualizar hasta cuatro comportamientos de la distribución en un mismo gráfico.</li> <li>4. Para el cálculo del “área bajo la curva” se requiere especificar las áreas sombreadas a graficar, estas son: cola a la derecha, cola a la izquierda, ambas colas e intermedio. Posteriormente se requiere el tipo de distribución, parámetros y el acumulado.</li> </ol>
<b>Prioridad del requisito</b> <input type="checkbox"/> Alta/Eencial <input checked="" type="checkbox"/> Media/Deseado <input type="checkbox"/> Baja/ Opcional

Fuente: Elaborado por el IMT

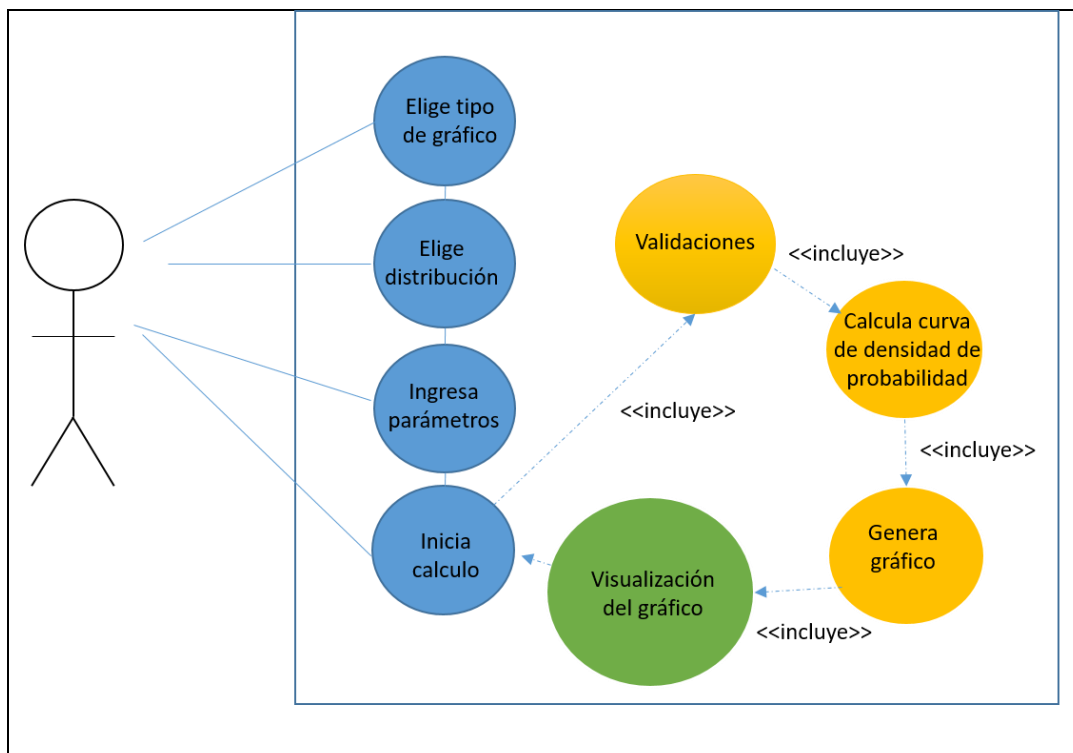


Figura 2.7 Diagrama de caso de uso de comparación de distribuciones.

## 2.7 Despliegue de nuevas funcionalidades de ABAL

A continuación, se presenta el despliegue de las nuevas funcionalidades de ABAL. Se inicia con las vistas de las distribuciones: logística, Logarítmica y Gumbel, posteriormente se muestran las figuras de gráficos Q-Q y P-P. Finalmente se muestra un prototipo de los gráficos de comparación de distribuciones.

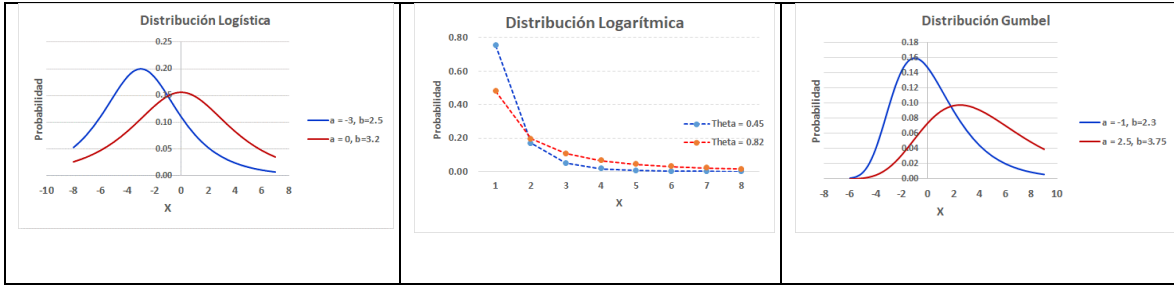


Figura 2.8 Vista de la distribución Logística, Logarítmica y Gumbel.

En la figura 2.9 se muestra una vista de las opciones de gráficas Q-Q y P-P en la sección de “otros gráficos”. Estos gráficos están disponibles para las distribuciones continuas.

The screenshot shows a software interface with two main sections: "GENERA HISTOGRAMA" and "OTROS GRÁFICOS".

- GENERA HISTOGRAMA:** Contains two text input fields: "Título del histograma" and "Título en eje X". Below these fields are two blue buttons: "Calcular prueba Ji-Cuadrada" and "Calcular prueba K-S".
- OTROS GRÁFICOS:** Contains two checkboxes: "Gráfica Q-Q" and "Gráfica P-P", both of which are currently unchecked.

Figura 2.9 Vista de opciones de gráficas Q-Q y P-P.

En la figura 2.10 se muestra un ejemplo de las gráficas Q-Q y P-P para una distribución Uniforme. Se observa (para todas las gráficas de este tipo) al costado derecho, un conjunto de herramientas para manipular los gráficos, de las cuales destaca la opción: “Save”, ícono en forma de disco, el cual, permite descargar los gráficos en la carpeta de “Descargas” de Windows, para su uso posterior.

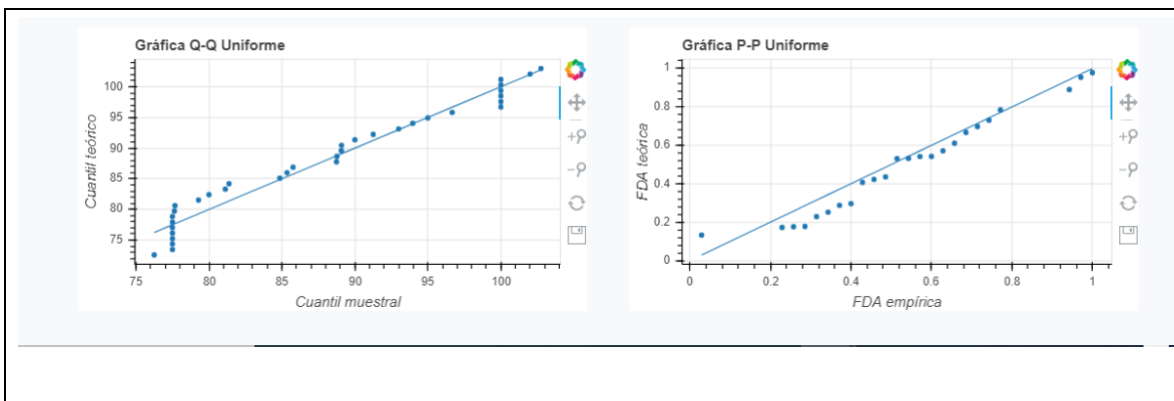


Figura 2.10 Vista de las gráficas Q-Q y P-P para una distribución Uniforme.

En la figura 2.11 se muestra una vista de las gráficas Q-Q y P-P para una distribución Normal.

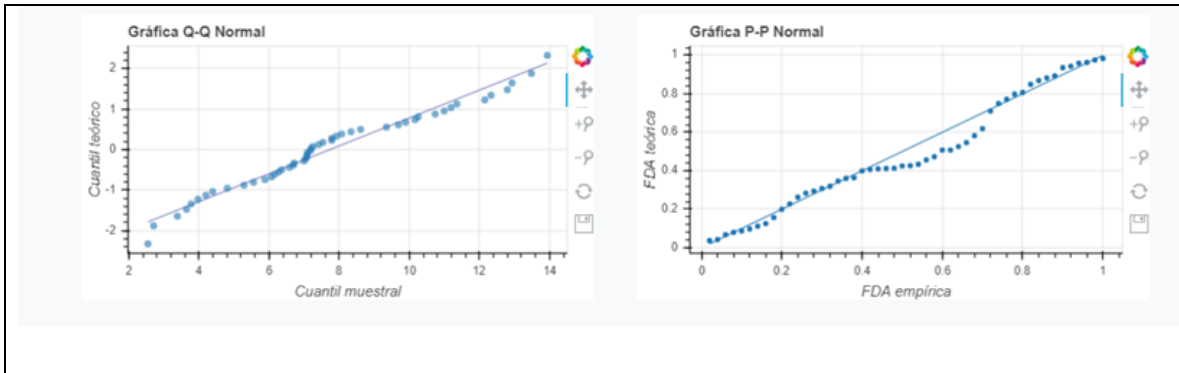


Figura 2.11 Vista de las gráficas Q-Q y P-P para una distribución Normal.

En la figura 2.12 se muestra una vista de las gráficas Q-Q y P-P para una distribución Exponencial.

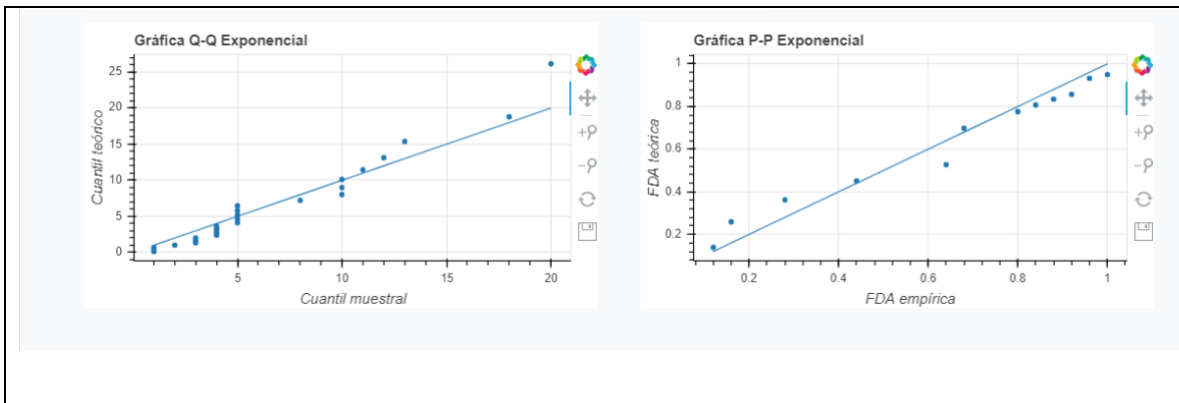


Figura 2.12 Vista de las gráficas Q-Q y P-P para una distribución Exponencial.

En la figura 2.13 se muestra una vista de las gráficas Q-Q y P-P para una distribución Lognormal.

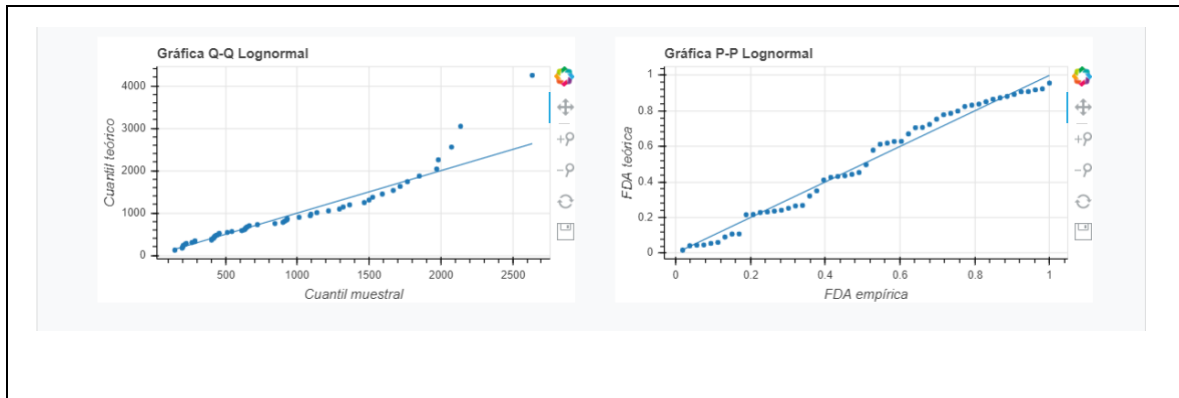
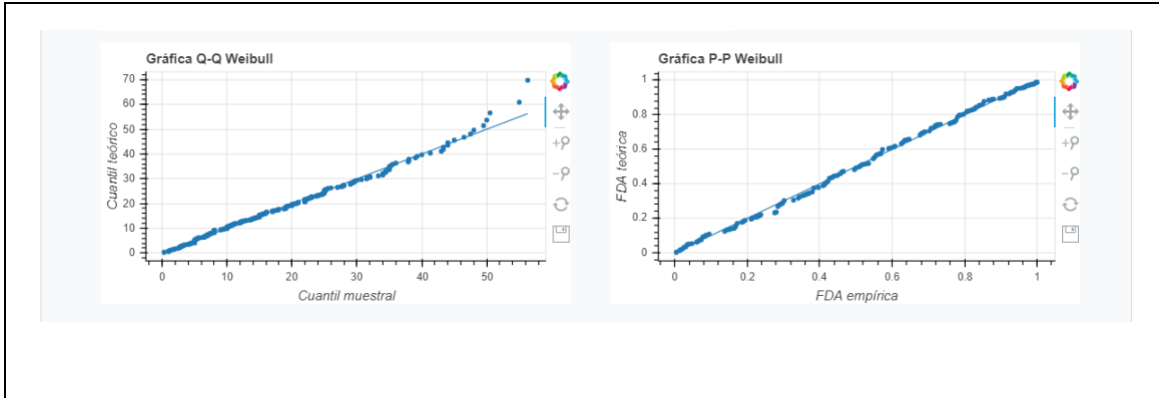


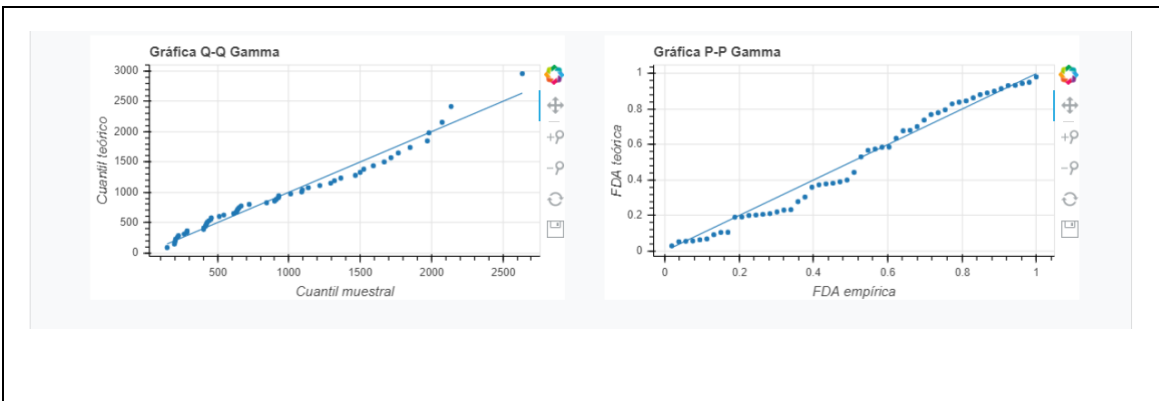
Figura 2.13 Vista de las gráficas Q-Q y P-P para una distribución Lognormal.

En la figura 2.14 se muestra una vista de las gráficas Q-Q y P-P para una distribución Weibull.



**Figura 2.14 Vista de las gráficas Q-Q y P-P para una distribución Weibull.**

En la figura 2.15 se muestra una vista de las gráficas Q-Q y P-P para una distribución Gamma.



**Figura 2.15 Vista de las gráficas Q-Q y P-P para una distribución Gamma.**

En la figura 2.16 se muestra una vista del prototipo de las gráficas de comparación de distribuciones. Conforme se realicen más pruebas en producción, este diseño podría variar ligeramente.

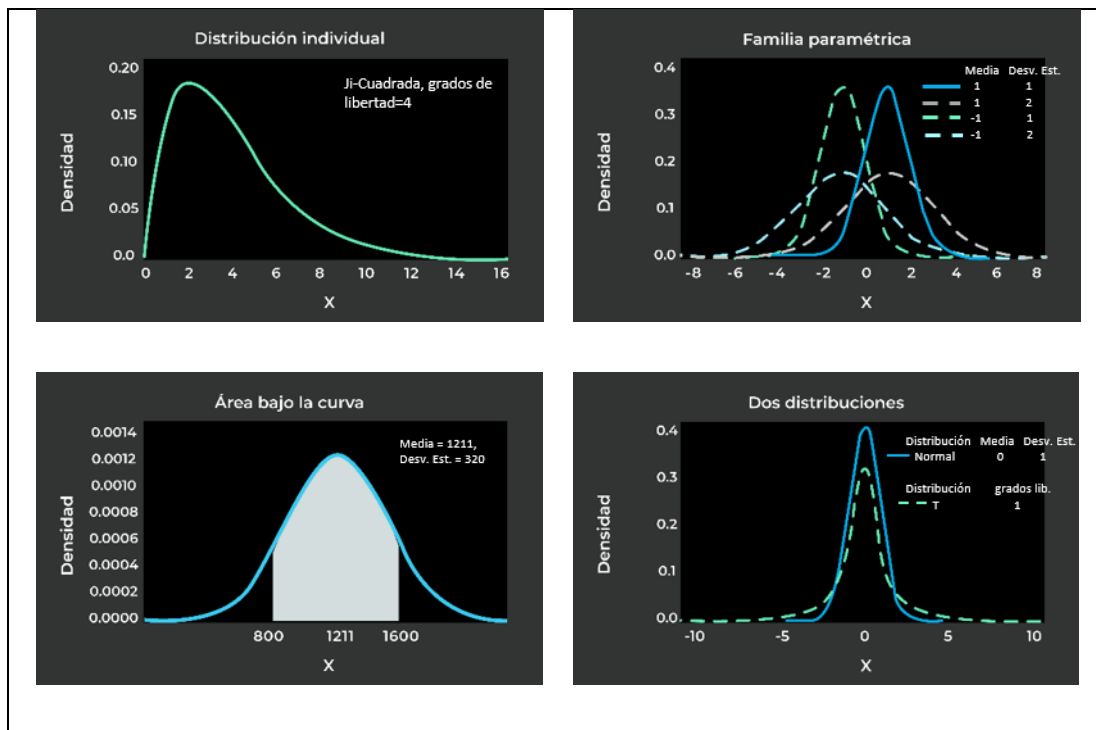


Figura 2.16 Vista de las gráficas de comparación de distribuciones.

### 3. Otras distribuciones de probabilidad

---

En la primera etapa de la aplicación ABAL se desarrollaron las pruebas de bondad de ajuste de seis distribuciones continuas: Uniforme, Normal, Exponencial, Lognormal, Weibull y Gamma, así como de tres distribuciones discretas: Binomial, Geométrica (en sus dos versiones, núm. de fallas previas al éxito y núm. de ensayos para el éxito) y Poisson.

En esta segunda etapa se agregan tres distribuciones más: la distribución discreta Logarítmica, y las distribuciones continuas Logística y Gumbel.

#### 3.1 Distribución Logarítmica

La distribución Logarítmica es una distribución discreta que toma valores positivos enteros a partir de 1, y que presenta colas alargadas hacia la derecha. En la literatura es común que se presente como una derivación basada en el desarrollo de la serie de Maclaurin para el logaritmo natural de la variable  $(1 - x)$ :

$$-\ln(1 - x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots; \text{ para } -1 \leq x < 1$$

Con un parámetro  $0 < \theta < 1$  y la igualdad de la serie de Maclaurin anterior, se verifica la siguiente ecuación, que se interpreta como la suma de probabilidades discretas que define a la distribución logarítmica:

$$\sum_{k=1}^{\infty} \frac{-1}{\ln(1 - \theta)} \frac{\theta^k}{k} = 1$$

Por lo que la función de probabilidad de la variable aleatoria Logarítmica  $X$  es:

$$P[X = k] = \frac{-1}{\ln(1 - \theta)} \frac{\theta^k}{k}; \quad k = 1, 2, \dots, \infty; \quad 0 < \theta < 1$$

Esta distribución puede modelar conteo de eventos de interés, donde el resultado son números enteros positivos a partir de 1. Por comodidad en las expresiones suele escribirse:  $a = \frac{-1}{\ln(1 - \theta)}$  con lo que la función de probabilidad resulta:

$$P[X = k] = a \frac{\theta^k}{k}; \quad k = 1, 2, \dots, \infty; \quad 0 < \theta < 1; \quad a = \frac{-1}{\ln(1 - \theta)}$$

La media y la varianza de la distribución Logarítmica son como sigue.

$$\text{Media} = \frac{a\theta}{(1 - \theta)}$$

$$\text{Varianza} = \frac{a\theta(1 - a\theta)}{(1 - \theta)^2}$$

Aunque la motivación de la distribución Logarítmica parece puramente matemática, ha resultado útil en situaciones de conteo donde puede obtenerse cualquier número mayor a cero. Algunos casos de uso exitoso de esta distribución son los siguientes.

- Andreassen, D.C. (1986). Logra un buen ajuste a una distribución logarítmica al analizar frecuencias de accidentes en Australia.
- Deni, S.D. and Jemain, A.A. (2009). Combinan una distribución logarítmica con una geométrica en un estudio de secuencias de días lluviosos y días secos en Malasia.
- TutorVista.com, un servicio de tutoría universitaria en línea, reporta como aplicaciones de la distribución logarítmica, la modelación de reclamaciones en aseguradoras y el conteo del número de productos que adquieren los consumidores en cierto periodo de observación.

La figura 3.1 muestra dos ejemplos de esta distribución, para  $\theta$  igual a 0.5 y 0.75.

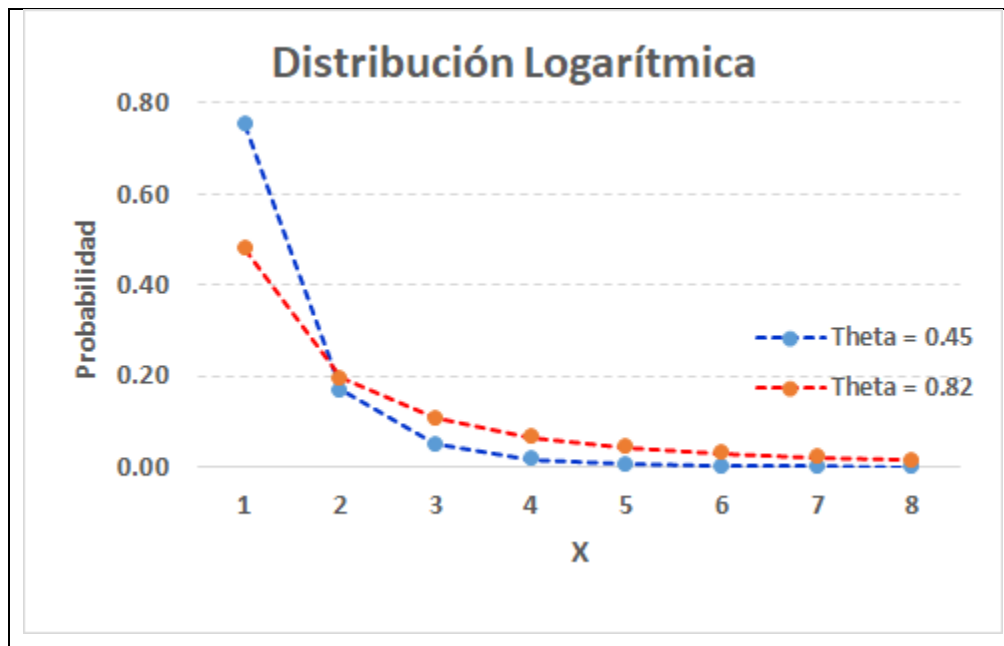


Figura 3.1 Función de probabilidad Logarítmica.



## 3.2 Distribución Logística

Esta distribución es llamada así debido a que su función de distribución acumulada (FDA) resulta ser una *curva logística*, la cual es usada típicamente en la modelación de crecimiento de poblaciones.

Una de sus aplicaciones más conocida es la *regresión logística o logit*, para estudios de elecciones discretas de bienes o servicios de usuarios que enfrentan alternativas de decisión, como por ejemplo elegir entre viaje en autobús o en tren, bajo diferentes ofertas de precio y tiempo.

El rango de valores de una variable aleatoria  $X$  logística de parámetros  $a$  (de localización),  $b$  (de escala) y su función de densidad de probabilidad (fdp) son:

$$f(x) = \frac{e^{-\left(\frac{x-a}{b}\right)}}{b \left[1 + e^{-\left(\frac{x-a}{b}\right)}\right]^2}$$

$$-\infty < x < \infty; \quad -\infty < a < \infty; \quad b > 0$$

La media y la varianza de la distribución Logística son como sigue.

$$\text{Media} = a$$

$$\text{Varianza} = \frac{b^2 \pi^2}{3}$$

La distribución logística es simétrica respecto al parámetro de localización  $a$ , y en algunas circunstancias puede reemplazar a la función Normal, y al igual que ésta, cuando se “normaliza” la variable aleatoria original, se obtiene la densidad  $(0, 1)$ , es decir, si:

$$X \sim \text{Logistica}(a, b), \text{ entonces la v. a } Y = \frac{X - a}{b} \sim \text{Logistica}(0, 1)$$

Diversos autores han reportado el uso de la distribución Logística en trabajos de investigación en varios campos. Algunos ejemplos son:

- Van Beek, P. (1978) utiliza una distribución Logística para modelar una política de revisión continua para control de inventarios.
- Rajkai, K. et al. (1996) modelan con una distribución Logística características de retención de agua a partir de datos de densidad de suelos en Suecia.
- Cooray, K. (2005) analiza datos de tiempos de vida introduciendo una distribución Logística modificada que le permite un manejo mejorado de colas muy largas en la distribución de los datos.

- Krishnamoorthy, K. (2006) describe el uso de la Logística en comparaciones con la Weibull para estudios de velocidades de viento.
- Bowling, S.R. et al (2009), desarrollan un método de aproximación a la función de distribución acumulada  $\Phi(x)$  de la Normal, usando una distribución Logística.

La figura 3.2 muestra dos ejemplos de esta distribución.

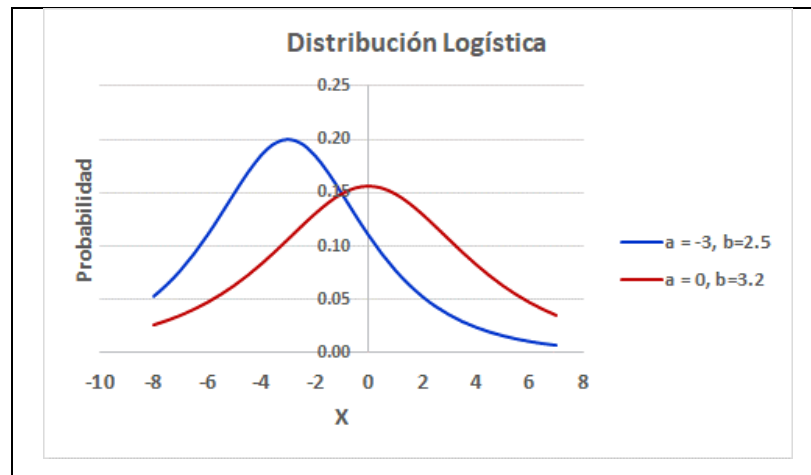


Figura 3.2 Densidad de probabilidad Logística.

### 3.3 Distribución Gumbel

Esta distribución, propuesta por E. J. Gumbel en 1935; se desarrolló para estudiar los valores más grandes (extremos) en muestras de datos y originalmente se aplicó para estimar niveles de inundaciones (Forbes, C. et al, 2011). También se conoce como Distribución de valores extremos Tipo I.

El rango de valores de una variable aleatoria  $X$  Gumbel de parámetros  $a$  (de localización),  $b$  (de escala) y su función de densidad de probabilidad (fdp) son como sigue:

$$f(x) = \frac{1}{b} e^{-\left[\frac{x-a}{b} + e^{-\left(\frac{x-a}{b}\right)}\right]}$$

$$-\infty < x < \infty; \quad -\infty < a < \infty; \quad b > 0$$

La media y la varianza correspondientes son como sigue, donde  $\gamma$  es la constante de Euler-Mascheroni  $\approx 0.5772156649$ :

$$\text{Media} = a + \gamma b$$

$$\text{Varianza} = \frac{b^2 \pi^2}{6}$$

La distribución Gumbel ha resultado apropiada como modelo probabilístico en situaciones donde interesa estimar probabilidades de que la v.a. de interés alcance o rebase valores de referencia dados (p. ej. en hidrología para niveles de lluvia o de sequía; en el ambiente industrial la estimación del mínimo tiempo de ocurrencia a una falla en un equipo).

Algunos trabajos de investigación que ejemplifican el uso de esta distribución son los siguientes.

- Ónoz, B. and Bayazit, M. (1995), reportan el uso de la distribución Gumbel como la de mejor desempeño para modelar niveles máximos de inundaciones con datos de varias regiones del mundo.
- Xu, Y.L. (1995) utiliza una distribución Gumbel para modelar las características de fatiga por presiones de viento en un edificio de una universidad de Texas.
- Jánosikova, L. and Slavik, M. (2014) modelan la llegada de pasajeros a paradas de autobús de transporte público en la República Eslovaca utilizando una distribución Gumbel.
- Rehman, K; Burton, P.W. and Weatherill, G.A. (2018) utilizan una distribución Gumbel y el método de simulación Monte Carlo en el análisis de riesgo sísmico en Pakistán.

La figura 3.3 muestra un par de ejemplos de esta distribución.

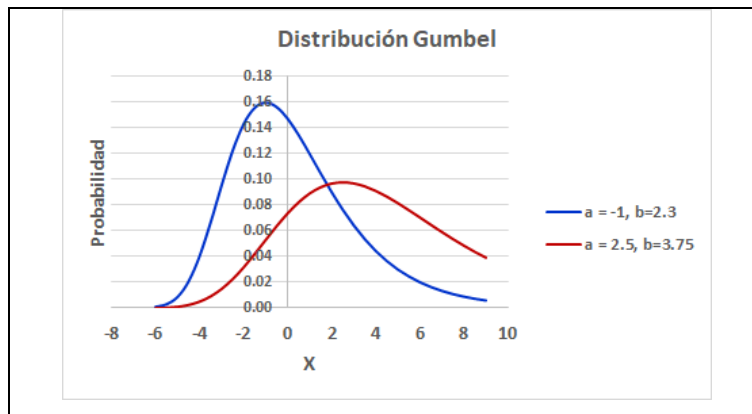


Figura 3.3 Densidad de probabilidad Gumbel.



## 4. La estimación de máxima verosimilitud

---

El proceso de estimación de parámetros de una distribución de probabilidad es el primer paso en la determinación de la bondad de ajuste de cualquier muestra de datos que se ingrese a la aplicación ABAL. Los dos métodos de estimación de parámetros usados en la primera versión de ABAL fueron: a) el método de momentos, y b) el método de máxima verosimilitud.

El método de momentos para estimar los parámetros de una distribución se basa en igualar los momentos de la distribución teórica propuesta para los datos, con los momentos muestrales calculados de los datos que se están ajustando. Este método es sencillo de usar, y en general permite estimar los parámetros por procedimientos algebraicos, pero puede dar estimaciones fuera de rango de los parámetros teóricos en algunos casos cuando se manejan muestras pequeñas.

El método de máxima verosimilitud estima los parámetros de una distribución buscando el valor numérico de los parámetros que maximizan la probabilidad de haber obtenido la muestra particular que se está analizando.

Este método produce resultados mucho más precisos que el método de momentos, pero exige un tratamiento matemático más elaborado. Si la distribución de probabilidad a probar es diferenciable, se pueden aplicar los métodos clásicos de optimización, pero en muchos casos el sistema de ecuaciones que resulta para encontrar los puntos críticos es demasiado complicado y se requieren métodos numéricos para encontrar las soluciones.

En este capítulo se muestra el desarrollo de estimadores de máxima verosimilitud para las distribuciones que en la primera versión de ABAL se manejaron con el método de momentos, y para las nuevas distribuciones en este trabajo de ampliación de funcionalidades de ABAL. Las distribuciones consideradas son:

- a) Logarítmica (discreta)
- b) Weibull (continua)
- c) Logística (continua)
- d) Gumbel (continua)
- c) Gamma (continua).

## 4.1 Distribución Logarítmica

Esta distribución discreta tiene la siguiente función de probabilidad.

$$P[X = k] = \frac{-1}{\ln(1 - \theta)} \frac{\theta^k}{k}; \quad k = 1, 2, \dots, \infty; \quad 0 < \theta < 1$$

Dada una muestra  $x_1, x_2, \dots, x_n$  con media muestral  $\bar{X}$ , el estimador de máxima verosimilitud  $\hat{\theta}$  para la Logarítmica cumple con la ecuación (Krishnamoorthy, 2006):

$$\hat{\theta} = \frac{\bar{X} \text{Ln}(1 - \hat{\theta})}{\bar{X} \text{Ln}(1 - \hat{\theta}) - 1} \quad \dots (4.1)$$

El único dato requerido para el cálculo es la media muestral  $\bar{X}$ , sin embargo la ecuación (4.1) no calcula  $\hat{\theta}$  de manera explícita, y no es sencillo despejar el valor de  $\hat{\theta}$ , por lo que es necesario utilizar métodos numéricos para la solución. Si se tiene la media  $\bar{X}$ , se podría aproximar el valor de  $\hat{\theta}$ , considerando la diferencia  $d(\hat{\theta})$ :

$$d(\hat{\theta}) = \hat{\theta} - \frac{\bar{X} \text{Ln}(1 - \hat{\theta})}{\bar{X} \text{Ln}(1 - \hat{\theta}) - 1} \quad \dots (4.1a)$$

Al graficar  $\hat{\theta}$  contra  $d(\hat{\theta})$  puede ubicarse el valor que hace  $d(\hat{\theta}) = 0$ , donde la curva cruza el eje horizontal. La figura 4.1 muestra el corte con el eje horizontal entre 0.7 y 0.8. como primera aproximación a  $\hat{\theta}$  para una media  $\bar{X} = 2.15$

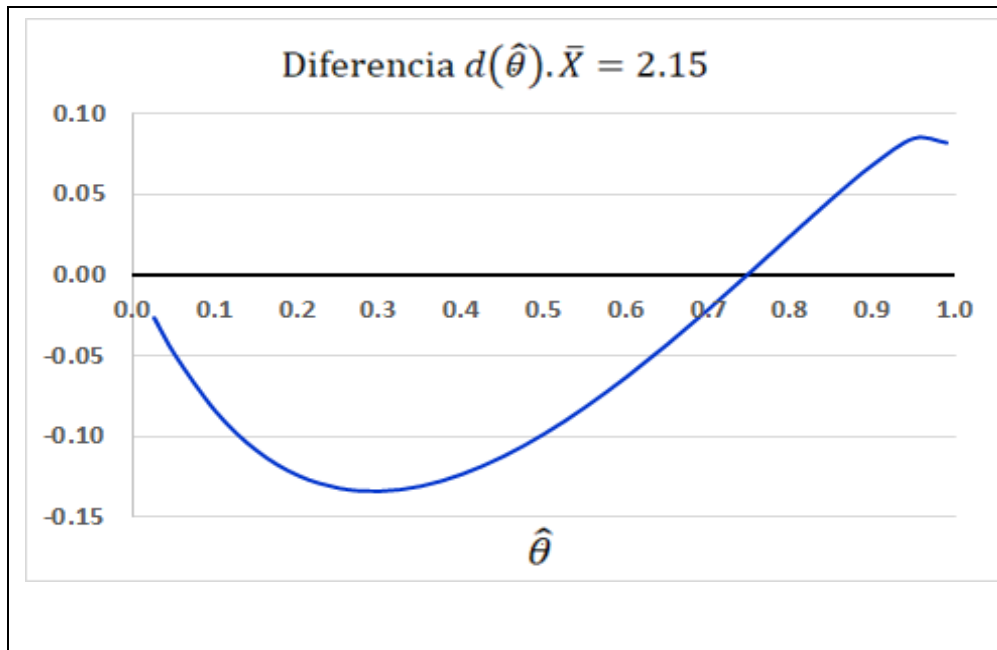


Figura 4.1 Aproximación gráfica al parámetro  $\hat{\theta}$ .

Para afinar la estimación de  $\hat{\theta}$  se utiliza el complemento Solver de Excel, en la opción de optimización no-lineal, usando como punto de búsqueda inicial 0.7, que es límite inferior del intervalo donde se encuentra el cruce con el eje horizontal en la figura 4.1. La figura 4.2 muestra los valores inicial y final de la solución con Solver, donde la celda Objetivo contiene la ecuación (4.1a).

TABLA INICIAL DE SOLVER			TABLA FINAL DE SOLVER		
$\bar{X}$	$\hat{\theta}$	Objetivo	$\bar{X}$	$\hat{\theta}$	Objetivo
2.15	0.7	-0.02134	2.15	0.74732453	0.00000

Figura 4.2 Estimación del parámetro  $\hat{\theta}$  con Solver de Excel.

## 4.2 Distribución Weibull

La distribución Weibull tiene dos parámetros:  $\lambda$  de forma y  $k$  de escala. El rango de valores de  $x$  y su función de densidad de probabilidad (fdp) son:

$$f(x) = \frac{\lambda}{k} \left(\frac{x}{k}\right)^{\lambda-1} e^{-(x/k)^\lambda}; \quad x > 0; \quad \lambda, k > 0$$

En la primer versión de ABAL se estiman los parámetros con una rutina de regresión lineal basada en linealizar la función de distribución acumulada con una transformación logarítmica; esta estimación se usa para los valores iniciales  $\lambda_0, k_0$  en la rutina de optimización.

Dada una muestra de datos  $x_1, x_2, \dots, x_n$  para la estimación MLE de los parámetros  $\lambda, k$ , debe maximizarse el logaritmo de la función de verosimilitud (Nielsen, 2011):

$$LL(\lambda, k) = n \ln(\lambda) - n\lambda \ln(k) + (\lambda - 1) \sum_{j=1}^n \ln(x_j) - \sum_{j=1}^n \left(\frac{x_j}{k}\right)^\lambda \quad \dots (4.2)$$

Los valores  $\hat{\lambda}, \hat{k}$  que maximizan la ecuación (4.2) son la estimación MLE requerida. Dado el carácter no-lineal de esta ecuación, la maximización puede hacerse con un algoritmo de gradiente generalizado (GRG), que está disponible en una rutina tipo Solver como la que ofrece Excel. El procedimiento de cálculo se ilustra en la hoja de Excel de la figura 4.3.

La muestra de datos es de tamaño 75, pero sólo se ven los primeros 10 datos. La estimación de valores iniciales  $\lambda_0, k_0$  se obtienen con la rutina de estimación para Weibull usada en ABAL 1.0, y aparece en la tabla marcada en amarillo.

[1]	[2]	[3]		[4]	[5]	[6]
Xj	Ln Xj	(Xj/k) <sup>λ</sup>	#Datos n	λ	k	OBJ
9.4	2.24071	0.32616	75	2.03800	16.28830	-256.58097
16.1	2.77882	0.97658				
11.3	2.42480	0.47465				
14.6	2.68102	0.80011				
19.3	2.96011	1.41307	<b>Valores iniciales con ABAL 1.0</b>			
10.9	2.38876	0.44103	<b>Parámetros λ= 2.0380, k= 16.2883</b>			
9	2.19722	0.29850	# datos= 75, gr. de libertad= 6,			
8	2.07944	0.23480	test Ji-Cuadrada= 6.9745,			
20.7	3.03013	1.62984	valor crítico (nivel signif. α=5%)= 12.5916			
24.8	3.21084	2.35554	valor-p=0.3232			

Figura 4.3 Estimación de parámetros Weibull λ,k con Solver de Excel.

Los elementos de la tabla están organizados como sigue.

Los valores iniciales  $\lambda_0, k_0$  generados con ABAL 1.0 están en las columnas [4] y [5].

La columna [1] tiene los datos  $x_j$  (son 75; sólo se ven los primeros 10). La columna [2] tiene el valor  $\ln(x_j)$ . Para facilitar los cálculos, la columna [3] contiene el valor  $(x_j/k)^\lambda$  con  $\lambda, k$  en columnas [4] y [5].

El objetivo a maximizar en la columna [6] tiene la ecuación (4.2), con  $\lambda, k$  en columnas [4] y [5]. Las sumatorias se obtienen con las columnas [2] y [3], y el valor de  $n = 75$  que se muestra en la tabla.

Con la ecuación (4.2) en la columna [6] (el objetivo) se corre Solver para maximizar la casilla OBJ y se encuentran los estimadores MLE para la distribución Weibull:

La solución de Solver se muestra en la figura 4.4.  $\hat{\lambda} = 1.94912, \hat{k} = 16.70240$

Figura 4.4 Solución MLE de parámetros Weibull λ,k con Solver de Excel.



### 4.3 Distribución Logística

La distribución Logística se describe con la siguiente función de densidad de probabilidad (fdp) de dos parámetros  $a, b$ :

$$f(x) = \frac{e^{-\left(\frac{x-a}{b}\right)}}{b \left[1 + e^{-\left(\frac{x-a}{b}\right)}\right]^2}$$

$$-\infty < x < \infty; \quad -\infty < a < \infty; \quad b > 0$$

Dada una muestra de datos  $x_1, x_2, \dots, x_n$  para la estimación MLE de los parámetros  $a, b$ , el logaritmo de la función de verosimilitud que debe maximizarse es como sigue (Zaiontz, C., 2019):

$$LL(a, b) = \frac{n}{b} a - \frac{n}{b} \bar{X} - 2 \sum_{j=1}^n \ln \left[ 1 + \exp \left( - \left( \frac{x_j - a}{b} \right) \right) \right] - n \ln(b) \quad \dots (4.3)$$

Donde  $\bar{X} = \frac{1}{n} \sum_{j=1}^n x_j$

La maximización se hace con la opción de optimización no-lineal de Solver; los valores iniciales  $a_0, b_0$  se obtienen con base en el método de momentos como sigue.

$$a_0 = \bar{X} = \frac{1}{n} \sum_{j=1}^n x_j; \quad b_0 = \frac{\sqrt{3}}{\pi} \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{X})^2}$$

La tabla 4.5 muestra la tabla de cálculos en Excel.

[1]	[2]			[3]	[4]	[5]
Xk	Ln(1+exp-(Xk-a)/b)	#Datos n	Promedio Xk	a	b	OBJETIVO
-5.022	1.173364	45	-3.3814	-3.38135	2.04236	-122.57339
-4.739	1.079913					
-4.613	1.039482					
0.805	0.121149					
-1.196	0.294904	<b>Valores iniciales de paráms.</b>				
-3.726	0.781164	<b>a0</b>	<b>b0</b>			
-1.699	0.363747	<b>-3.38135</b>	<b>2.04236</b>			
-6.782	1.838255					
-8.965	2.796793					
-2.408	0.482901					

Figura 4.5 Estimación de parámetros a,b de la distribución Logística.

La columna [1] tiene los datos  $x_j$ .

La columna [2] tiene el valor  $\ln \left[ 1 + \exp \left( - \left( \frac{x_j - a}{b} \right) \right) \right]$  usando  $a, b$  de las columnas [3] y [4].

El objetivo a maximizar en la columna [5] tiene la ecuación (4.3), usando  $a, b$  de las columnas [3] y [4].

La casilla G3 para el valor de  $b$  se restringe en el cuadro de Solver como  $> 0.0001$  para asegurar que el algoritmo no pueda ensayar el valor cero en sus rutinas.

Con la ecuación (4.3) en la casilla OBJETIVO, se corre Solver para maximizar el valor en esa casilla y se obtienen los estimadores MLE para la distribución Logística, como se muestra en la figura 4.6.

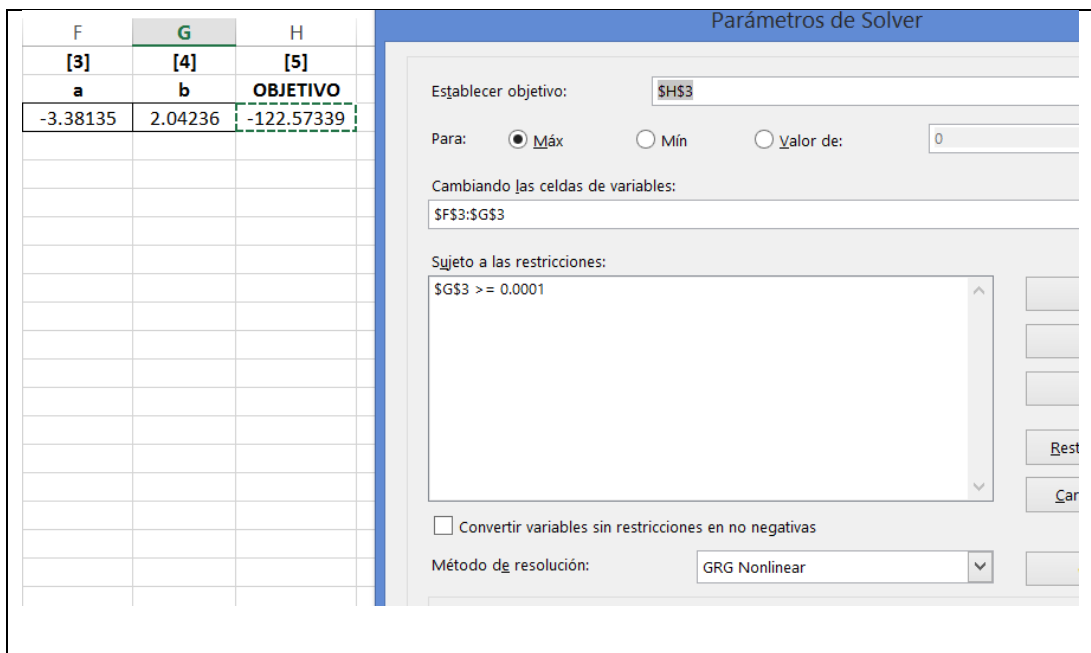


Figura 4.6 Parámetros de máxima verosimilitud  $a, b$  en distribución Logística.

La solución de Solver indica los estimadores MLE  $\hat{a} = -3.38135$ ,  $\hat{b} = 2.04236$

## 4.4 Distribución de Gumbel

La función de densidad de probabilidad y los parámetros de esta distribución son:

$$f(x) = \frac{1}{b} e^{-\left[ \frac{x-a}{b} + e^{-\left( \frac{x-a}{b} \right)} \right]}$$

$$-\infty < x < \infty; \quad -\infty < a < \infty; \quad b > 0$$

Dada una muestra de datos  $x_1, x_2, \dots, x_n$  el estimador de máxima verosimilitud  $b$  cumple la siguiente condición (Zaiontz, C., 2019):

$$b = \bar{x} - \frac{\sum_{j=1}^n x_j \exp\left(-\frac{x_j}{b}\right)}{\sum_{j=1}^n \exp\left(-\frac{x_j}{b}\right)} \dots (4.4)$$

El valor de  $a$  se estima a partir de  $b$  como:

$$a = -bLn \left[ \frac{1}{n} \sum_{j=1}^n \exp\left(-\frac{x_j}{b}\right) \right] \dots (4.5)$$

La dificultad para tratar la ecuación (4.4) es que al estar  $b$  en ambos lados de la ecuación, no se puede despejar para dejar el parámetro  $b$  de un solo lado.

Una manera de aproximar el valor de  $b$  es considerar la diferencia  $D$  de la ecuación (4.4.a) la cual se deriva de la ecuación (1):

$$D = b - \left( \bar{x} - \frac{\sum_{j=1}^n x_j \exp\left(-\frac{x_j}{b}\right)}{\sum_{j=1}^n \exp\left(-\frac{x_j}{b}\right)} \right) \dots (4.4a)$$

Probando valores de  $b$  en un cierto rango, p.ej. entre 1 y 3.5, se obtiene la gráfica de la figura 4.7 donde se aprecia que la curva corta al eje horizontal en algún punto entre 2.6 y 2.8.

Este procedimiento puede dar una guía para encontrar una buena aproximación para  $b$  sin embargo no es práctico, ya que el rango a elegir para graficar no es obvio a partir del conjunto de datos, además de que el grado de precisión logrado depende del incremento usado en los valores tentativos de  $b$ .

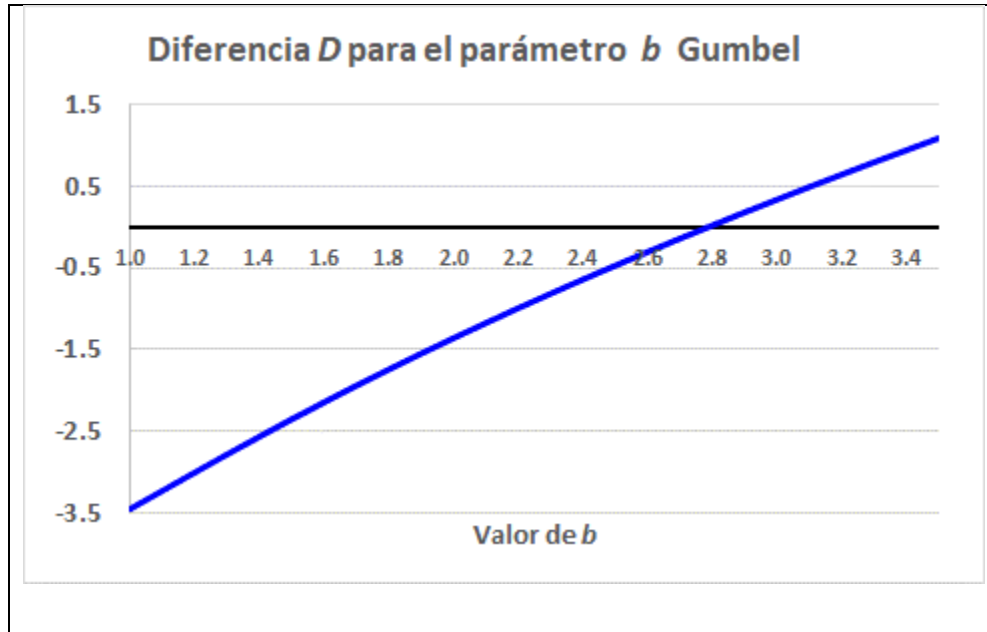


Figura 4.7 Aproximación gráfica al parámetro b Gumbel.

Una mejor manera para estimar  $b$  se puede usar Solver de Excel para obtener cero o el valor más cercano posible en la diferencia de la ecuación (4.4a). El valor inicial  $b_0$  para correr Solver se estima con el método de momentos como indica la ecuación (4.6), donde  $\bar{X}$  es la media muestral.

$$b_0 = \frac{\sqrt{6}}{\pi} \sqrt{\frac{\sum_{j=1}^n (x_j - \bar{X})^2}{n - 1}} \dots (4.6)$$

La figura 4.8 muestra la tabla de cálculos en Excel.

[1]	[2]	[3]			[4]	[5]
Xk	Exp (-Xk/b)	Xk*Exp(-Xk/b)	#Datos n	Xprom	b	OBJETIVO
-0.236	1.08742	-0.25638	45	1.37073	2.81337	0.0501312412
-1.328	1.60331	-2.12935				
1.205	0.65161	0.78518			a	
0.930	0.71843	0.66838			-0.23761	
-0.372	1.14125	-0.42422	Valor inicial b0			
1.930	0.50364	0.97186	2.81337			
1.998	0.49151	0.98217				
13.162	0.00930	0.12234				
0.701	0.77944	0.54641				
5.382	0.14762	0.79454				

Figura 4.8 Estimación de parámetros a,b de la distribución Gumbel.

La columna [1] contiene los datos  $x_j$ . La columna [2] contiene el valor  $\exp(-x_j/b)$ .

La columna [3] contiene el valor  $x_j \exp(-x_j/b)$ . La columna [4] contiene el valor  $b$  que cambia Solver para lograr cero en la casilla OBJETIVO.

La columna [5] contiene la diferencia  $D$  de la ecuación (4.4a) que se calcula con las sumatorias de las columnas [2] y [3] más el valor  $b$  en la columna [4].

El procedimiento inicia poniendo el valor inicial  $b_0$  calculado con la ecuación (4.6) en la casilla de la columna [4] y luego corriendo Solver con la condición de poner 0 en la casilla OBJETIVO. El valor de  $a$  se estima automáticamente con la ecuación (4.5) que aparece en anaranjado en la hoja de Excel de la figura 4.9.

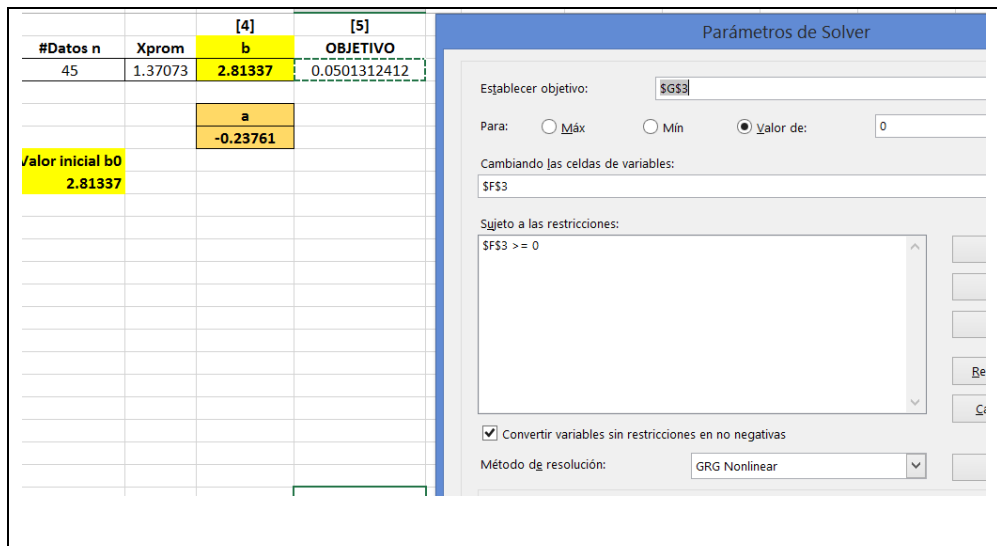


Figura 4.9 Parámetros de máxima verosimilitud a,b en distribución Gumbel.

La solución de Solver es:  $\hat{a} = -0.25056$ ,  $\hat{b} = 2.78208$ .

## 4.5 Distribución Gamma

La función de densidad de probabilidad y sus parámetros son:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)}$$

$\alpha$ : forma,  $\beta$ : escala;  $\alpha, \beta > 0$ ;  $x \geq 0$

El logaritmo de la función de máxima verosimilitud que hay que maximizar es como sigue:

$$LL(\alpha, \beta) = n \left[ (\alpha - 1) \overline{\ln(x_j)} - \frac{\bar{X}}{\beta} - \alpha \ln(\beta) - \ln(\Gamma(\alpha)) \right] \dots (4.7)$$

Donde  $\overline{\ln(x_j)}$  es el promedio de los logaritmos de los datos  $x_j$ , y  $\bar{X}$  es el promedio de los  $x_j$ .

Se sabe que el máximo de (4.7) ocurre si  $\beta = \frac{\bar{X}}{\alpha}$  (Zaointz, C., 2019); así que con ese valor para  $\beta$ , la estimación de  $\alpha$  se hace con Solver maximizando la ecuación (4.7).

El valor inicial  $\alpha_0$  se estima con el método de momentos con la ecuación (4.8) :

$$\alpha_0 = \frac{m_1'^2}{m_2' - m_1'^2} \dots (4.8)$$

La figura 4.10 muestra la hoja de Excel para los cálculos.

[1]	[2]	[3]	[4]	[5]	[6]	[7]	
Xk	Ln(Xk)	m'1	m'2	#Datos n	$\alpha$	$\beta$	OBJETIVO
1312.8	7.180	827.432	835032.811	41	4.55246	181.75486	-300.4898011
734.3	6.599						
427.4	6.058	Valor inicial					
744.3	6.612	$\alpha_0$					
686.9	6.532	4.55246					
954.1	6.861						
1147.5	7.045						
550.2	6.310						
701.6	6.553						
596.5	6.391						

Figura 4.10 Estimación de parámetros  $\alpha, \beta$  de la distribución Gamma.

La columna [1] contiene los datos  $x_j$ . La columna [2] contiene los valores  $\ln(x_j)$ .

Las columnas [3] y [4] contienen los momentos muestrales para calcular el valor inicial  $\alpha_0$  con la ecuación (4.8).

La columna [5] contiene el valor  $\alpha$  para la estimación con Solver. La columna [6] contiene el valor  $\beta$  calculado a partir de  $\alpha$  con la ecuación  $\beta = \frac{\bar{X}}{\alpha}$ .

La columna [7] contiene la ecuación (4.7) que es lo que maximiza Solver, y usa los valores  $\alpha, \beta$  de las columnas [5] y [6]. El planteamiento con Solver es:

[5]	[6]	[7]
$\alpha$	$\beta$	OBJETIVO
4.55246	181.75486	-300.4898011

Establecer objetivo:

Para:  Máx  Mín  Valor de:

Cambiando las celdas de variables:

Sujeto a las restricciones:

\$F\$3 >= 0

Convertir variables sin restricciones en no negativas

Método de resolución:

Figura 4.11 Estimación de parámetros  $\alpha, \beta$  de la distribución Gamma.

La solución que se obtiene de Solver es:

[5]	[6]	[7]
$\alpha$	$\beta$	OBJETIVO
4.27780	193.42476	-300.4462076

Figura 4.12 Solución de Solver.





## 5. Ejemplos de uso de ABAL con datos reales

---

La versión ABAL 0.1 ofrece en su menú una colección de ejemplos de bondad de ajuste usando datos reales de diversas fuentes, que se resolvieron con las rutinas que ofrece ABAL. Estos ejemplos muestran casos de las distribuciones que aparecen en el menú, como son: Uniforme, Normal, Exponencial, Lognormal, etc.

En este capítulo se muestran algunos ejemplos más, con las nuevas distribuciones que se han agregado al menú: la distribución Logarítmica del tipo discreto, y las distribuciones continuas Logística y Gumbel. Los ejemplos mostrados usan datos reales de distintas fuentes.

### 5.1 Ejemplo de la distribución Logarítmica

La distribución Logarítmica es del tipo discreto y los posibles valores que puede tomar son los números enteros positivos: 1, 2, 3, ... Esta característica la hace apropiada para modelar conteo de eventos aleatorios que sólo pueden tomar esos valores.

Los datos de la tabla 5.1 son número de tripulantes de vehículos utilitarios registrados en la Estación 246 "Playas de Tijuana" en las encuestas origen-destino 2010 del Estudio Estadístico del Autotransporte Nacional (EECAN) del IMT. El ajuste a una distribución Logarítmica se muestra enseguida.

**Tabla 5.1 Núm. de tripulantes de vehículos utilitarios. Estación 246 "Playas de Tijuana" EECAN 2010.**

1	2	4	2	1	2	1
1	1	2	4	2	1	1
2	1	2	1	1	1	1
1	3	1	1	2	1	2
1	2	1	1	2	1	1
1	2	1	1	1	1	1
2	4	1	2	1	1	1
2	1	3	1	1	2	2

Fuente: Estudios realizados por el IMT

La muestra es de tamaño 56; el máximo es 4 y su promedio es  $\bar{X} = 1.5357$ .

Este promedio se usa en el procedimiento descrito en la sección 4.1 y se obtiene la estimación del parámetro  $\hat{\theta}$ . El resultado de Solver se muestra enseguida

$\theta$	Objetivo
<b>0.55263</b>	<b>0.000000033</b>

Con esta estimación  $\hat{\theta} = 0.55263$  se realizó una prueba Ji-cuadrada de bondad de ajuste, con el resultado que se muestra en la figura 5.1.

Interv	Prob	FO	FE	JiCuad
1	0.68704	34	38.47409	0.52028
2	0.18984	17	10.63091	3.81579
3	0.06994	2	3.91662	0.93791
4	0.02899	3	1.62332	1.16751
>= 5	0.02420	0	1.35507	1.35507
		<b>56</b>	<b>56</b>	<b>7.7966</b>
$\hat{\theta}$	<b>0.55263</b>		<b>g.l.</b>	<b>V.Crít</b>
			4	9.48773
				<b>valor-p</b>
				0.09932

Figura 5.1 Prueba Ji-cuadrada para distrib. Logarítmica.

El valor de la prueba Ji-cuadrada es 7.7966, menor al valor crítico para 4 grados de libertad que es 9.48773. En la tabla se aprecia también el valor-p de esta prueba que es 9.932%, que supera el 5% de significancia de la prueba.

Este resultado indica que la variable número de tripulantes en vehículos utilitarios se distribuye como una Logarítmica de parámetro 0.55623.

La figura 5.2 muestra el histograma de los datos en términos de probabilidad empírica (barras color naranja) comparado al ajuste de la prueba Ji-cuadrada (en línea de puntos azul), que corresponde a la distribución teórica de los datos.

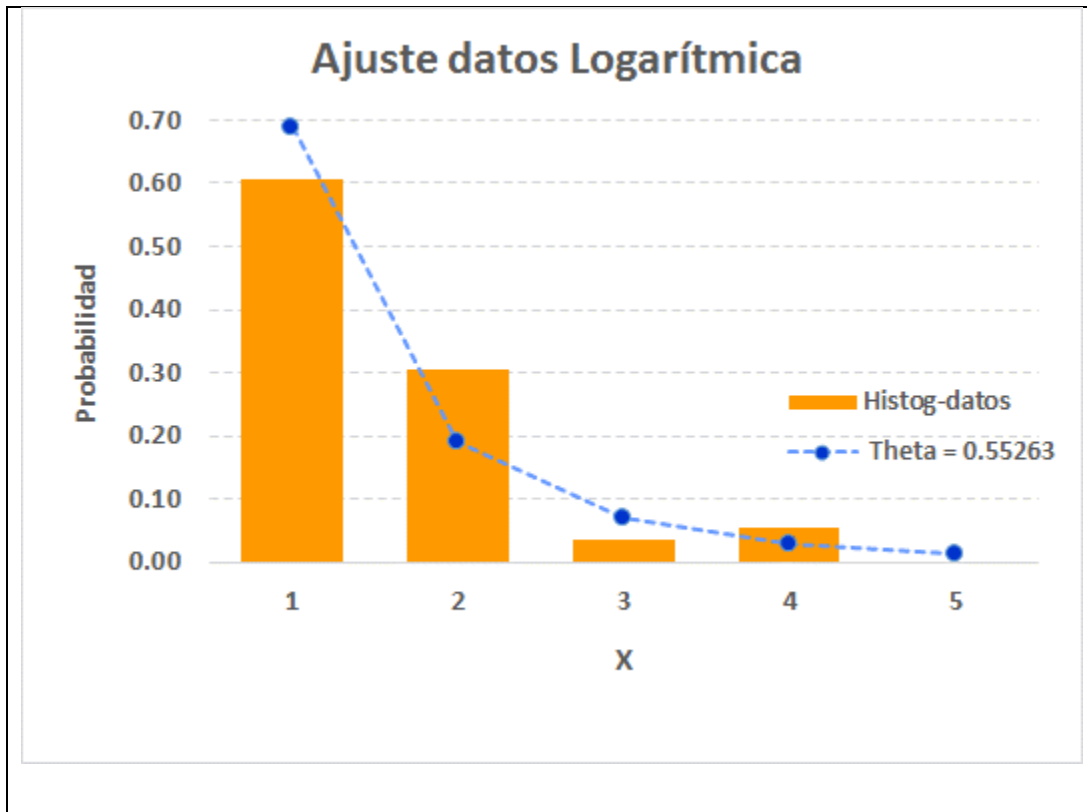


Figura 5.2 Prueba Histograma de datos y ajuste de Logarítmica  $\hat{\theta} = 0.55263$ .

## 5.2 Ejemplo de la distribución Logística

La distribución Logística es del tipo continuo, es simétrica alrededor de su parámetro  $\alpha$ , su dominio de valores es el eje real  $(-\infty, \infty)$  y en ocasiones puede usarse en lugar de la Normal.

Los datos de la tabla 5.2 son envíos mensuales de algodón (ton) exportado a Estados Unidos en 2019 por el cruce de Nuevo Laredo, registrados en el sistema Transborder Freight Data del Bureau of Transportations Statistics norteamericano (<https://www.bts.gov/transborder>).

Tabla 5.2 Exportación de algodón a EUA por Laredo, 2019 (tons).

816.947	683.282	755.913
627.784	541.986	543.690
627.123	702.175	477.182
484.088	799.890	594.595

Fuente: Estudios realizados por el IMT

El ajuste de estos datos a una distribución Logística se desarrolla enseguida.

Los valores iniciales  $a_0, b_0$  se obtienen como sigue.

$$a_0 = \bar{X} = \frac{1}{n} \sum_{j=1}^n x_j = 637.88792$$

$$b_0 = \frac{\sqrt{3}}{\pi} \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{X})^2} = 63.96661$$

Con estos valores se aplica el procedimiento de la sección 4.3, y se obtiene la estimación de máxima verosimilitud de los parámetros. El resultado de Solver se muestra en la figura 5.3

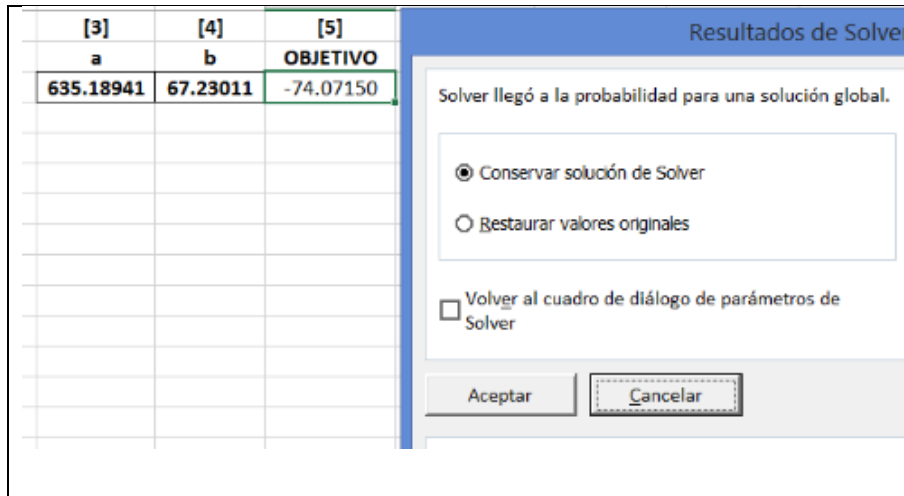


Figura 5.3 Estimación de parámetros a, b de la Logística.

La estimación de máxima verosimilitud indica los parámetros para la distribución Logística de los embarques de algodón como  $\hat{a} = 635.18941, \hat{b} = 67.23011$ .

La gráfica Q-Q que se muestra en la figura 5.4 indica que los datos ajustan razonablemente a una distribución logística con los parámetros estimados, con la mayoría de puntos cercanos a la recta de igualdad  $Y = X$ .

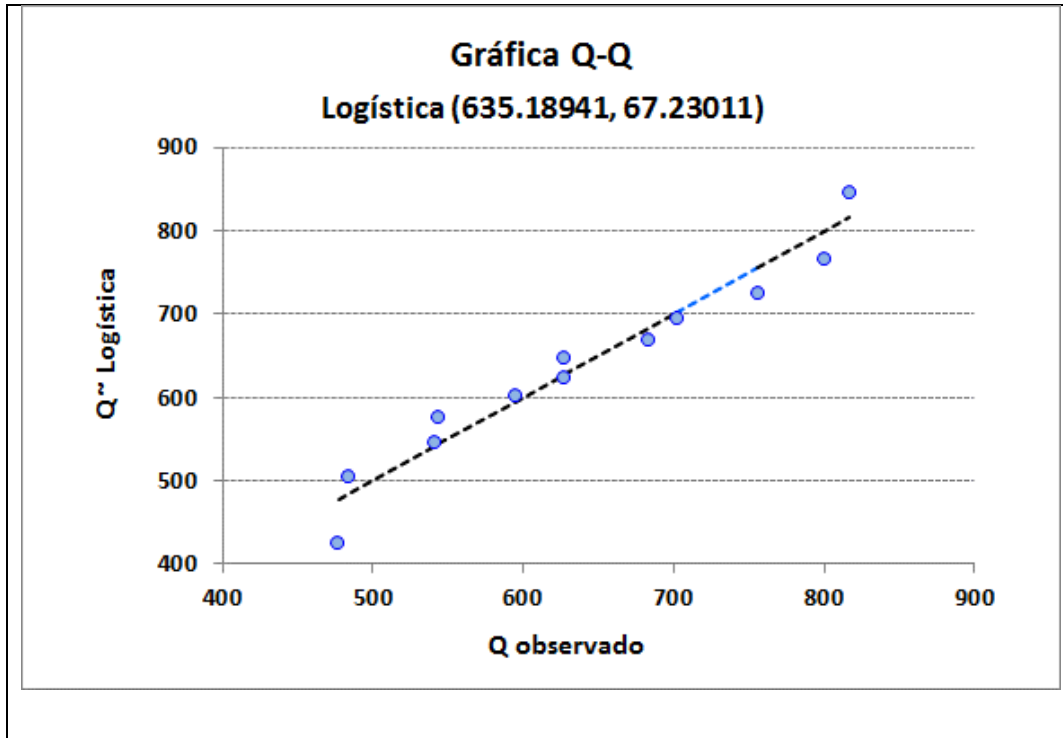


Figura 5.4 Gráfica Q-Q de datos de exportación de algodón ajustados a una distribución Logística.

### 5.3 Ejemplo de la distribución Gumbel

La distribución Gumbel o distribución de Valores Extremos es continua, con valores en el eje real  $(-\infty, \infty)$  y tiene dos parámetros  $a, b$  con  $a$  en el eje real y  $b$  positivo.

La tabla 5.3 muestra el valor de daños materiales (miles de dólares) en accidentes registrados en carreteras federales a nivel estatal en 2018 en el Anuario estadístico de accidentes del IMT (Cuevas A.C., Mayoral, E. F. y Mendoza A., 2018).

Tabla 5.3 Daños materiales (miles de dólares) por accidentes en carreteras federales por estado en 2018.

778.68	415.11	618.36	1617.62
1545.92	993.02	1812.89	1601.37
613.26	3498.75	4711.31	1756.29
778.92	2408.81	1029.41	1972.04
1061.40	2246.00	2550.28	1339.93
522.16	2769.22	2116.42	6257.96
2202.65	3719.10	879.96	703.84
1374.34	2790.62	2273.27	2045.60

Fuente: Estudios realizados por el IMT

Estos datos muestran buen ajuste a una distribución Gumbel, como se explica enseguida. El valor inicial del parámetro  $b$  se calcula como sigue:

$$b_0 = \frac{\sqrt{6}}{\pi} \sqrt{\frac{\sum_{j=1}^n (x_j - \bar{X})^2}{n-1}} = 1000.03855$$

Donde el promedio de los datos  $\bar{X} = 1906.39094$ .

Con este valor, se usa el procedimiento de la sección 4.4 para obtener la estimación de máxima verosimilitud de los parámetros  $a, b$  Gumbel. La solución con Solver se muestra en la figura 5.5.

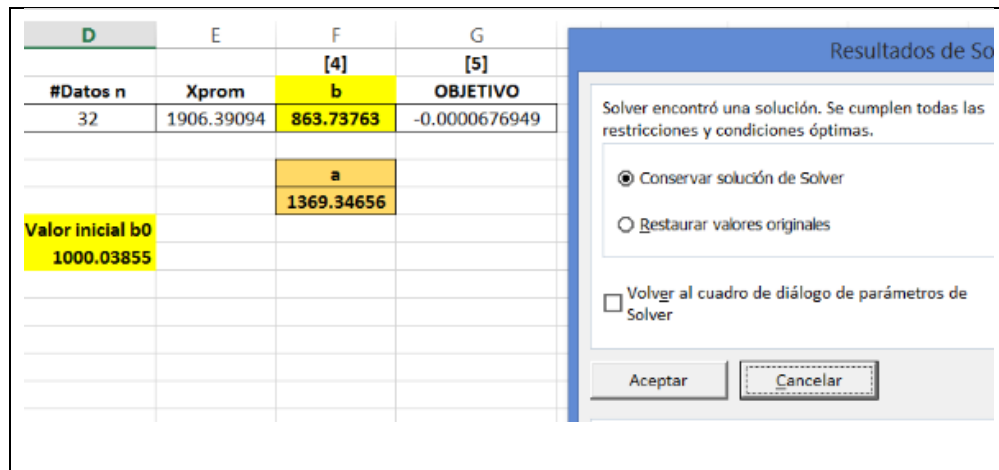


Figura 5.5 Estimación de parámetros  $a, b$  de la Gumbel.

El resultado de la figura 5.5 indica un ajuste de los datos a una distribución Gumbel de parámetros  $a = 1369.34656$ ,  $b = 863.73763$ .

La figura 5.6 muestra la gráfica Q-Q para este ajuste, donde se aprecia buena cercanía a la recta de igualdad  $Y = X$ , excepto por algunos valores en la cola superior, donde hay más distancia respecto a la recta.

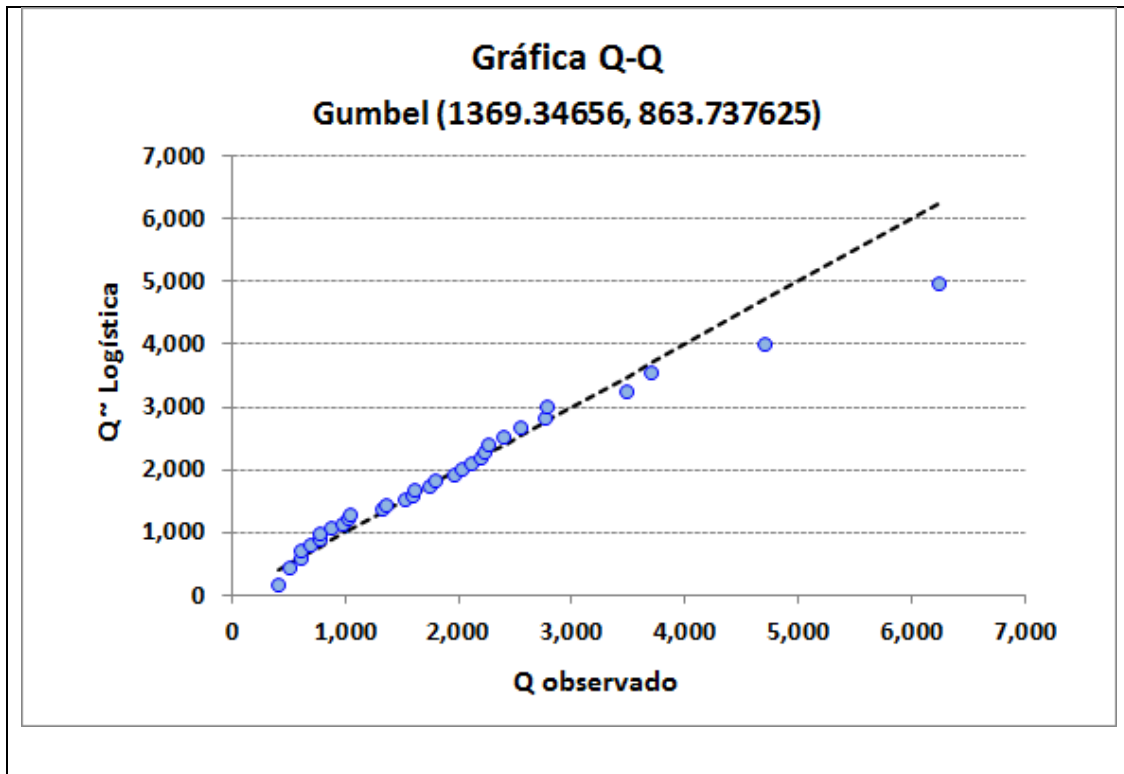


Figura 5.6 Gráfica Q-Q de datos de daños materiales en accidentes carreteros ajustados a una distribución Gumbel.

Los ejemplos mostrados indican que en cada caso se tiene una buena estimación de parámetros para cada una de las distribuciones consideradas. Para un resultado más formal, se pueden correr las respectivas rutinas de ABAL para dar el resultado de estimación completo con algunas de las pruebas estadísticas disponibles.





## 6. Resultados y conclusiones

---

Con la ampliación de funcionalidad de la aplicación para bondad de ajuste en línea, que se presenta en la versión ABAL 1.0, se dispone ya de un menú con ocho distribuciones continuas y cinco distribuciones discretas para ajustar datos de modelos probabilísticos de uso común en ingeniería y en el campo del transporte.

La adición de la capacidad de generar gráficas Q-Q y gráficas P-P de los datos modelados, añade la posibilidad de una visualización del ajuste de los datos como un complemento gráfico que facilita la interpretación de los resultados permitiendo que los usuarios tengan reportes de resultados más completos.

Por otra parte, el módulo gráfico que genera densidades de probabilidad que el usuario identifica con los parámetros de su interés, así como la capacidad de tener gráficas múltiples para comparar densidades distintas y sus coberturas de probabilidad entre los límites que el usuario indique, son de utilidad para decidir el modelo de probabilidad que mejor se adapte al contexto de modelación del usuario. Así, cuando en un conjunto de datos el usuario se encuentre con que dos densidades distintas, p. ej. la Weibull y la Logística ajustan satisfactoriamente con sus datos, el usuario puede decidir con base en su discusión teórica, o en los antecedentes de la disciplina en que trabaja (transporte, ingeniería civil, economía, etc.) cuál de las densidades aceptables se adapta mejor a su marco de referencia.

La adición de los ejemplos de ajuste de datos con casos reales para las nuevas distribuciones que se ofrecen en ABAL 1.0, brinda a los usuarios una colección más amplia de datos reales para que pueda realizar sus pruebas y detectar oportunidades de uso de las distribuciones que se muestran en los ejemplos.

### **Líneas de trabajo futuro**

Con el desarrollo e integración de las nuevas funcionalidades de la aplicación para bondad de ajuste en línea, la versión ABAL 1.0 ofrece a los usuarios una herramienta estadística con mayores alcances.

Sin embargo, durante su desarrollo y la revisión de fuentes bibliográficas, surgieron nuevas ideas que podrían brindar al usuario más opciones de cálculo. Por ejemplo, mostrar nuevos estilos de gráficas más atractivas que faciliten la comparación entre diferentes representaciones visuales, y así, brindar una influencia sustancial en los conocimientos que se pueden extraer de la visualización.

Algunos ejemplos de ello son:

- Especificar la cantidad de intervalos de clase o categorías.

- En algunos casos podría ser útil "encoger" las barras ligeramente para enfatizar la naturaleza categórica del eje. Ver figura 6.1.

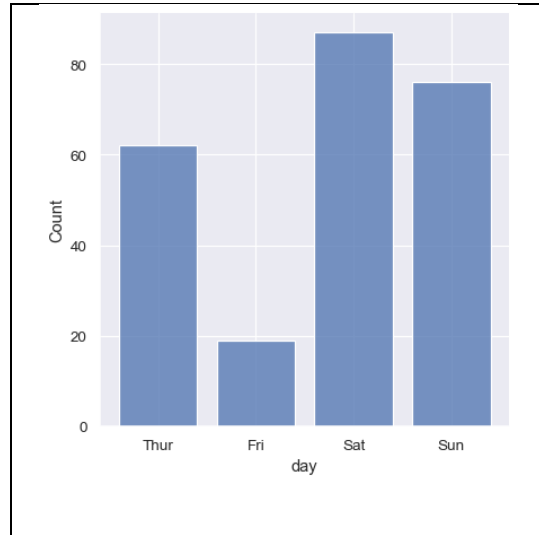


Figura 6.1 Enfatizar la categoría del eje.

- De forma predeterminada, los diferentes histogramas están "superpuestos" uno encima del otro y, en algunos casos, pueden ser difíciles de distinguir, (Waskom, M., 2012-2020). Una opción es tener una casilla disponible para activar el cambio de la representación visual del histograma a un diagrama de "pasos". Ver figura 6.2.

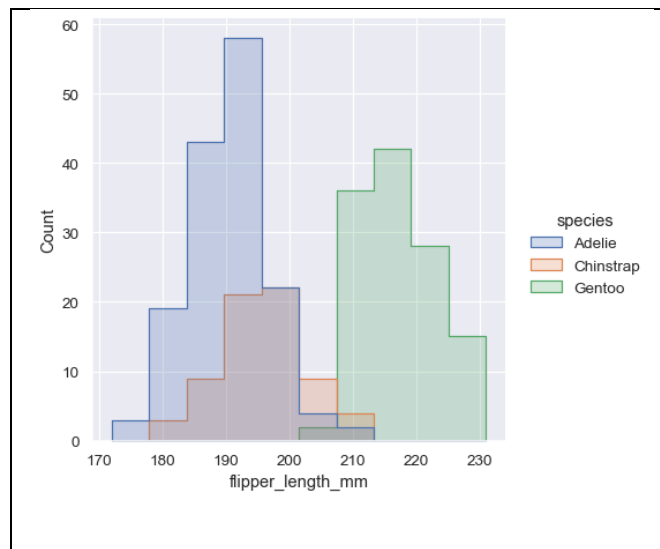


Figura 6.2 Diagrama de pasos.

Durante el desarrollo de esta fase, se identificó D3.js, una librería JavaScript para manipular documentos basados en datos. Esta es una buena opción para desarrollar gráficos interactivos, particularmente beneficioso para la comparación de distribuciones, debido a que trabaja directamente sobre las capacidades que los navegadores Web ofrecen.

Con la revisión de ejemplos en la Web, se observó que la ejecución de gráficos con D3 es muy rápida, y además, admite grandes conjuntos de datos y comportamientos dinámicos. En este sentido, se pretende explorar más a fondo esta librería y si resulta viable, se implementará más adelante en ABAL.

Por otro lado, es importante mencionar que, una vez que se actualicen las nuevas funcionalidades en ABAL y sean usadas por un gran número de usuarios en la Web, se espera continuar recibiendo su valiosa retroalimentación. Se evaluarán las que resulten pertinentes y se procederá a realizar los ajustes necesarios. De esta forma se mejorará continuamente la funcionalidad de ABAL. Los últimos cambios serán visibles en la sección de ayuda de la aplicación.

Es importante mencionar que, con las actualizaciones antes mencionadas, ABAL pasará de la versión 0.1 a la 1.0. Conforme se desarrollen más funciones, estas versiones se incrementarán hasta lograr una versión más estable y completa pero sobre todo, con incrementos que sean útiles para los usuarios.



## Bibliografía

---

Andreassen, D.C. (1986). "Intersection accident frequencies". *Traffic Engineering + Control*. October 1986, pp. 514-517.

Balaban, J. (2018). Una suave introducción a la estimación de máxima verosimilitud. Towards data science. Recuperado de: <https://n9.cl/cjkuv>

Bostock, M. (2020). Documentos basados en datos. D3. Recuperado de: <https://d3js.org/>

Bowling, S.R. et al. (2009). "A logistic approximation to the cumulative normal distribution". *Journal of Industrial Engineering and Management*. – 2(1): 114-127 – ISSN: 2013-0953. En: <http://www.jiem.org/index.php/jiem/article/viewFile/60/27>

Castilla, H. (2018). Resolviendo ejercicio de programación no lineal con Python – Investigación de operaciones. Medium. Recuperado de: <https://n9.cl/b2ekw>

Cooray, K. (2005). *Analyzing lifetime data with long-tailed skewed distribution: the logistic-sinh family*. Statistical Modelling Society SMS. En: <http://journals.sagepub.com/doi/abs/10.1191/1471082X05st0990a?journalCode=s mja>.

Deni, S.D. and Jemain, A.A. (2009). "Mixed log series geometric distribution for sequences of dry days". *Atmospheric Research*. Vo. 92, Issue 2, April 2009, Pages 236-243.

Desarrolladores de pyOpt. (2008-2014). Optimizadores. pyOpt. Recuperado de: <http://www.pyopt.org/reference/optimizers.html>

Equipo de desarrollo de SymPy. (2020). Solucionadores. SymPy. Recuperado de: <https://docs.sympy.org/latest/modules/solvers/solvers.html>

Evans, R. (2018). Estimación de máxima verosimilitud. Quant Econ Notas. Recuperado de: <https://n9.cl/efyb8>

Forbes, C., Evans, M., Hastings, N. and Peacock, B. (2011). *Statistical Distributions*. 4<sup>th</sup> Edition. John Wiley & Sons. USA

Jánosikova, L. and Slavik, M. (2014). "Modelling passenger's arrivals at public transport stops". *European Transport\ Trasporti Europei*. Issue 56, paper No. 7. En: [http://www.istiee.org/te/papers/N56/P07\\_56\\_12\\_2014.pdf](http://www.istiee.org/te/papers/N56/P07_56_12_2014.pdf)

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. Chapman & Hall/CRC. USA.

Nielsen, Mark A. (2011). *Parameter Estimation for the Tw-Parameter Weibull Distribution*. Nielsen, All Theses and Dissertations. 2509. Brigham Young University. Disponible en: <https://scholarsarchive.byu.edu/etd/2509>

Önoz, B. and Bayazit, M. (1995). "Best.fit distribution of largest available flood samples". *Journal of Hydrology*. 167 (1-4): 195-208. En: [https://www.researchgate.net/publication/222213660\\_Best-fit\\_distribution\\_of\\_largest\\_available\\_flood\\_samples](https://www.researchgate.net/publication/222213660_Best-fit_distribution_of_largest_available_flood_samples)

Perktold, J. Seabold, S. Taylor, J. (2009-2019). Estimación de máxima verosimilitud (modelos genéricos). Statsmodels v0.13.0.dev0 (+129). Recuperado de: [https://www.statsmodels.org/dev/examples/notebooks/generated/generic\\_mle.html](https://www.statsmodels.org/dev/examples/notebooks/generated/generic_mle.html)

Rajkai, K; Kabos, S; Van Genuchten, M.T. and Jansson, P.E. (1996). "Estimation of water-retention characteristics from the bulk density of Swedish soils". *Soil Science* 161(12):832-845. En: [https://www.researchgate.net/publication/234203956\\_Estimation\\_of\\_water-retention\\_characteristics\\_from\\_the\\_bulk\\_density\\_of\\_Swedish\\_soils](https://www.researchgate.net/publication/234203956_Estimation_of_water-retention_characteristics_from_the_bulk_density_of_Swedish_soils)

Rehman, K; Burton, P.W. and Weatherill, G.A. (2018). "Application of Gumbel I and Monte Carlo methods to assess seismic hazard in and around Pakistan". *Journal of Seismology* .Volume 22, Issue 3, pp 575-588. En: <https://link.springer.com/article/10.1007/s10950-017-9723-8>

TutorVista.com. (2018). *Logarithmic Distribution*. Disponible en: <https://math.tutorvista.com/statistics/logarithmic-distribution.html>

Van beek, P. (1978). "An application of the logistic density on a stochastic continuous review stock control model". *Unternehmensforschung Operations Research* 22(1). En: [https://www.researchgate.net/publication/40167730\\_An\\_application\\_of\\_the\\_logistic\\_density\\_on\\_a\\_stochastic\\_continuous\\_review\\_stock\\_control\\_model](https://www.researchgate.net/publication/40167730_An_application_of_the_logistic_density_on_a_stochastic_continuous_review_stock_control_model)

Waskom, M. (2012-2020). Visualización de distribuciones de datos. Seaborn. Recuperado de: <https://seaborn.pydata.org/tutorial/distributions.html>

Xu, Y.L. (1995). "Model- and full-scale comparison of fatigue-related characteristics of wind pressures on the Texas Tech Building". *Journal of Wind Engineering and Industrial Aerodynamics*. 58(3):147-173. En: [https://www.researchgate.net/publication/222726298\\_Model-\\_and\\_full-scale\\_comparison\\_of\\_fatigue-related\\_characteristics\\_of\\_wind\\_pressures\\_on\\_the\\_Texas\\_Tech\\_Building](https://www.researchgate.net/publication/222726298_Model-_and_full-scale_comparison_of_fatigue-related_characteristics_of_wind_pressures_on_the_Texas_Tech_Building)

Zaiontz, Charles. (2019). *Distribution Fitting via Maximum Likelihood*. Real Statistics Using Excel. Disponible en: <https://www.real-statistics.com/distribution-fitting/distribution-fitting-via-maximum-likelihood/>



# COMUNICACIONES

SECRETARÍA DE COMUNICACIONES Y TRANSPORTES



**Km 12+000 Carretera Estatal 431 "El Colorado Galindo"**  
Parque Tecnológico San Fandila, Mpio. Pedro Escobedo,  
Querétaro, México. C.P. 76703  
Tel: +52 (442) 216 97 77 ext. 2610  
Fax: +52 (442) 216 9671

[publicaciones@imt.mx](mailto:publicaciones@imt.mx)

<http://www.imt.mx/>