



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE CIENCIAS

MODELAJE DEL RIESGO DE INCUMPLIMIENTO EN  
EL TIEMPO DE ARRIBO AL AEROPUERTO BENITO  
JUÁREZ DE LA CDMX MEDIANTE EL MODELO  
AGREGADO

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

DANIEL ALEJANDRO ROCHA LÓPEZ

TUTOR

ACT. GABRIELA STEPHANIA REVUELTAS HERNÁNDEZ



CIUDAD UNIVERSITARIA, CDMX, 2019



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

*A mi padre y madre por su apoyo incondicional, este logro es para ustedes.  
Aiza y Alexis, hicieron de la universidad la mejor de las experiencias.  
A mis amigos, por acompañarme en este camino.*

## 1. Datos del alumno.

Rocha  
López  
Daniel Alejandro  
55 31 06 09 39  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
312302643

## 2. Datos del asesor

Act  
Revueltas  
Hernández  
Gabriela Stephanía

## 3. Datos del sinodal 1

Dra  
Ruth Selene  
Fuentes  
García

## 4. Datos del sinodal 2

Dr  
Yuri  
Salazar  
Flores

## 5. Datos del sinodal 3

M en C  
Alejandro  
Santoyo  
Cano

## 6. Datos del sinodal 4

Act  
Edgar  
Días  
Ordóñez

## 7. Datos del trabajo escrito

Modelaje del riesgo de incumplimiento en el tiempo de arribo al aeropuerto Benito Juárez de la CDMX mediante el modelo agregado  
Páginas 88  
2019

# Índice general

<b>Introducción</b>	<b>1</b>
<b>Contexto Histórico</b>	<b>2</b>
<b>1. Ajustes de distribuciones</b>	<b>6</b>
1.1. Distribuciones discretas . . . . .	6
1.2. Distribuciones continuas . . . . .	11
1.3. Ajuste de distribuciones . . . . .	17
<b>2. Modelo de Mezclas y algoritmo EM</b>	<b>21</b>
2.1. Mezclas de distribuciones . . . . .	21
2.1.1. Mezclas normales . . . . .	23
2.2. Algoritmo EM . . . . .	24
<b>3. El modelo agregado y sus aproximaciones</b>	<b>32</b>
3.1. Modelo Individual . . . . .	33
3.2. Modelo Compuesto . . . . .	36
3.3. Aproximaciones . . . . .	42
3.3.1. Aproximación Normal . . . . .	42
3.3.2. Aproximación Log-normal . . . . .	44
3.3.3. Aproximación Gamma trasladada . . . . .	45
<b>4. Ajuste y selección del modelo</b>	<b>46</b>
4.1. Presentación de los datos . . . . .	46
4.2. Ajuste del modelo agregado . . . . .	52
4.3. Ajuste de aproximaciones . . . . .	64
4.4. Interpretación y selección del modelo . . . . .	68
4.5. Modelos finales y su simulación . . . . .	69
<b>Conclusiones</b>	<b>74</b>
<b>Bibliografía</b>	<b>77</b>
<b>Anexos</b>	<b>79</b>

# Introducción

La actuaría siempre ha estado y estará relacionada con el aseguramiento, puesto que la necesidad de proteger bienes materiales frente a riesgos que generan pérdidas fue lo que dio origen a la ciencia actuarial. Por ello, matemáticos y actuarios han desarrollado teorías y técnicas que les permitieran analizar de manera integral los diferentes aspectos que conforman los riesgos.

El desarrollo de la teoría estocástica y de los modelos agregados fue un punto de quiebre en la ciencia actuarial, tan cierta es esta aseveración que hoy en día las compañías aseguradoras la siguen utilizando en el desarrollo de sus productos. Sin embargo, gracias a la formación integral que tienen los actuarios, es que se ha implementado el modelaje matemático en el análisis de sectores no tradicionales.

En esta tesis se realizará el proceso de modelado de datos oficiales de incumplimiento de arribo al Aeropuerto Internacional de la Ciudad de México de junio a noviembre del 2016 mediante el uso del modelo de riesgos compuestos o agregados.

Derivado de los Decretos publicados en el Diario Oficial de la Federación el 26 de junio de 2017 por los que se reforman la Ley de Aviación Civil y de la Ley Federal de Protección al Consumidor, se estableció un sistema de compensaciones a los consumidores de servicios aeroportuarios en caso de cancelación o retraso de vuelos comerciales.

Sin embargo, estas compensaciones están determinadas por las mismas aerolíneas, por lo que las entidades regulatorias, en este caso la Secretaría de Comunicaciones y Transporte y la Procuraduría Federal del Consumidor, carecen de métricas con fundamentos matemáticos sólidos para determinar la virtud de dichas compensaciones.

Con ello en mente, surgió la idea de generar un modelo matemático que provea información para analizar, no sólo cuántos incumplimientos de llegadas puntuales para vuelos comerciales incurren las aerolíneas, si no también qué tanto tiempo representa cada arribo tardío, siendo esto un símil de la regulación existente.

Así, el objetivo principal de la presente tesis es generar dos modelos, uno para el riesgo de incumplimiento en el tiempo de arribo por aerolíneas nacionales y otro para el riesgo de incumplimiento en el tiempo de arribo por aerolíneas extranjeras, que permitan la obtención de información y métricas sobre la frecuencia y severidad de los riesgos, tal que, en un futuro, la construcción de los modelos que queda asentada en este trabajo, pueda ofrecer una base con mayor objetividad para la publicación de nuevas regulaciones complementarias para el sector aeroportuario en el país.

Para lograr dicho objetivo se deberán ajustar distribuciones significativas a la severidad y frecuencia de los riesgos ya sea mediante densidades conocidas o mezclas, complementando el análisis mediante el uso de tres métodos usuales de aproximación del modelo agregado. Posteriormente, para cada riesgo se seleccionarán las distribuciones o aproximaciones que permitan construir un modelo que describa cada riesgo de manera acertada y que conserve el principio de parsimonia.

Finalmente, se generarán ambos modelos agregados mediante las distribuciones seleccionadas y se analizarán comprobando su consistencia contra la simulación de los riesgos.

En los primeros tres capítulos de este trabajo de tesis, se desarrollarán los conocimientos teóricos que se requieren para generar e interpretar los modelos. Por tanto, en el primer capítulo se presentan algunas de las distribuciones discretas y continuas más comunes para la composición de modelos agregados, así como la forma de ajustarlas a los datos. Similarmente el segundo capítulo se desarrolla primero la construcción de mezclas de distribuciones y sus propiedades, presentando después el algoritmo EM para su ajuste.

El modelo agregado se presenta en el tercer capítulo, tratando no sólo sus propiedades y las de las variables que lo componen, sino también aproximaciones al mismo. La aplicación de la teoría sobre los datos recopilados se realiza en el cuarto y último capítulo, en el cual se muestran los ajustes, gráficas y métricas obtenidos durante el proceso de ajuste realizado mediante el software estadístico **R**.

# Contexto Histórico

En primer lugar se establecerá el contexto histórico en el que se crearon tanto el aeropuerto de la Ciudad de México, como las instituciones que hoy en día son responsables de la operación del mismo ya que, en palabras de Margaret MacMillan,

*“La historia, al dar contexto y ejemplos, ayuda cuando toca pensar en el mundo presente. Nos ayuda a formular interrogantes, y sin buenos interrogantes, es muy difícil pensar de una forma coherente. El conocimiento de la historia sugiere el tipo de información que se podría requerir para responder esas preguntas. La experiencia nos enseña cómo evaluar esa información.”<sup>1</sup>*

El Aeropuerto Internacional “Benito Juárez” de la Ciudad de México tiene más de 50 años de historia, en ese tiempo se convirtió en el aeropuerto número uno en México en términos de pasajeros transportados, siendo que en 2017 transportó un total de 44,732,418 pasajeros, un 7.2 % más que el año anterior.<sup>2</sup>

Su historia se remonta hasta inicios del siglo XX cuando en 1928 se inició la construcción de la ahora Terminal 1 del aeropuerto bajo la dirección de la Secretaría de Comunicaciones y Obras Públicas (SCOP) para que 11 años después, en 1939, se inaugurara el “Puerto Aéreo Central de la Ciudad de México”. Sin embargo, no fue sino hasta el 2 de diciembre de 1963 que el entonces Secretario de Comunicaciones y Transportes, Walter Buchanan emitió el dictamen que le denominaba oficialmente como “Aeropuerto Internacional de la Ciudad de México”.<sup>3</sup>

Como parte del programa de reestructuración de la red aeroportuaria nacional, el 1º de noviembre de 1998 el AICM se desincorporó de la red del organismo público descentralizado Aeropuertos y Servicios Auxiliares (ASA) para conformar el Grupo Aeroportuario de la Ciudad de México, que consistía de tres empresas: Aeropuerto Internacional de la Ciudad de México, S.A. de C.V., Servicios Aeroportuarios de la Ciudad de México, S.A. de C.V. y Grupo Aeroportuario de la Ciudad de México,

---

<sup>1</sup>Margaret Olwen MacMillan, *The Uses and Abuses of History* (2008)

<sup>2</sup>Estadísticas y Movimiento Operacional, consulta 18/05/2018, fuente: <https://www.aicm.com.mx/estadisticas>

<sup>3</sup>Antecedentes Históricos, consulta 18/05/2018, fuente: <https://www.aicm.com.mx/aicm/acerca-del-aicm/breve-historia>



S.A. de C.V., todas ellas entidades que operan en calidad de empresas de participación estatal mayoritaria, y que originalmente operaban el aeropuerto en espera de su privatización.

En octubre de 2001, el entonces presidente Vicente Fox Quesada anunció que el gobierno federal construiría un aeropuerto alterno al de la Ciudad de México, en la zona del Lago de Texcoco, sin embargo el jueves 1 de agosto del 2002, mediante un comunicado la Secretaría de Comunicaciones y Transportes anunció la cancelación del proyecto finalizando 9 meses de protestas del Frente de Pueblos en Defensa de la Tierra (FPDT)<sup>4</sup>.

A raíz de estos eventos, el Gobierno Federal tomó la decisión de ampliar la terminal aérea existente en lugar de construir el nuevo aeropuerto. El programa de ampliación y remodelación del aeropuerto a su máxima capacidad incluía la construcción de una segunda terminal, lo que permitiría al aeropuerto incrementar su capacidad de atención de 24 a más de 32 millones de pasajeros nacionales e internacionales manejando aproximadamente el 43 % de las operaciones aéreas del Aeropuerto Internacional de la Ciudad de México.

El 24 de noviembre de 2006 se publica mediante acuerdo en el Diario Oficial de la Federación, el cambio de denominación del AICM a Aeropuerto Internacional Benito Juárez de la Ciudad de México (AIBJCM) y poco más de un año después, el 26 de marzo de 2008, se inaugura oficialmente la segunda terminal siendo presidente el Lic. Felipe Calderón Hinojosa.

A causa del resurgimiento de la propuesta de descentralizar el tráfico aéreo que transita por la Ciudad de México, la población mexicana renovó su interés en la estructura, organización y eficiencia (o deficiencia) del sector aeroportuario, en particular en la región centro del país.

La propuesta consistía en la construcción del Nuevo Aeropuerto de la Ciudad de México (NAIM) en los terrenos de la zona federal del ex Lago de Texcoco bajo los planos diseñados por el arquitecto mexicano Fernando Romero en colaboración con el arquitecto británico Norman Foster que consistían en un aeropuerto con dos terminales y 6 pistas de aterrizaje extendiéndose sobre una superficie aproximada de 5,000 hectáreas haciéndolo el aeropuerto más grande de latinoamérica.<sup>5</sup>

A pesar de la polémica que envolvió al sector a raíz de las elecciones presidenciales de 2018, no se puede negar que en la actualidad, la aeronáutica es un sector en crecimiento a nivel mundial, impulsando el desarrollo económico y la generación de

---

<sup>4</sup>La redacción, 1 agosto 2002, "Derrotan los ejidatarios a Fox: Cancela el aeropuerto en Texcoco", Proceso

<sup>5</sup>7 datos sobre el nuevo Aeropuerto de la Ciudad de México, consulta 20/05/2018, fuente: <https://aristeguinoticias.com/0309/mexico/7-datos-sobre-el-nuevo-aeropuerto-de-la-ciudad-de-mexico/>

empleos tanto fijos como temporales, tal como lo reporta Airports Council International.<sup>6</sup> Pues al día de hoy, el AICM es uno de los aeropuertos más importantes de Latinoamérica y el más grande de los 22 aeropuertos existentes dentro del territorio nacional y continúa apegándose a su misión de *‘satisfacer plenamente las necesidades de los usuarios del aeropuerto, a través de la prestación de servicios que consoliden el desarrollo, crecimiento y evolución del mismo’*.<sup>7</sup>

Según los datos presentados por la misma asociación, este sector económico generó en 2017 un ingreso global por cerca de 161.3 billones<sup>8</sup> de dólares americanos (USD) y un incremento anual promedio del 7.6 % respecto al año anterior.

En el reporte *Annual World Airport Traffic Report (WATR)* con cifras recopiladas hasta el 1 de enero de 2018, se ubica al Aeropuerto Internacional de la Ciudad de México en la posición 36 de los aeropuertos más transitados del mundo como se muestra en la siguiente tabla.

Rangp	País	Aeropuerto	Pasajeros
1	United States	Hartsfield-Jackson Atlanta International Airport	103,902,992
2	China	Beijing Capital International Airport	95,786,442
3	United A. Emirates	Dubai International Airport	88,242,099
4	United States	Los Angeles International Airport	84,557,968
5	United States	O'Hare International Airport	79,828,183
...	...	...	...
36	Mexico	Aeropuerto Internacional de la Ciudad de México	44,732,418

Tabla 1: Tráfico aéreo internacional, clasificado por pasajeros

Entre sus proyecciones al año 2040, la asociación Airports Council International incluye a México entre el Top 10 de países de los que se espera el mayor crecimiento a nivel mundial en transporte aéreo de pasajeros entre 2017 y 2040.

<sup>6</sup>Airport Statistics and Data Centre, consulta 20/05/2018, fuente: <https://aci.aero/data-centre/>

<sup>7</sup>Misión, visión y valores, consulta 20/05/2018, fuente: <https://www.aicm.com.mx/aicm/acerca-del-aicm/mision-vision-y-valores>

<sup>8</sup>161,300,000,000

# Capítulo 1

## Ajustes de distribuciones

En este capítulo se presentan las características básicas de algunas de las distribuciones discretas y continuas comúnmente utilizadas en modelos de riesgos y finalmente teoría básica relacionada a las pruebas de hipótesis para el ajuste de distribuciones.

### 1.1. Distribuciones discretas

Dentro del contexto del seguro, la frecuencia nos permite contabilizar el número de reclamaciones ocurridas, pero en un aspecto más amplio la frecuencia es una variable aleatoria que muestra cuántas veces se materializa el riesgo; por su misma naturaleza esta variable es asociada a una distribución discreta con probabilidad positiva en los enteros no negativos.

A continuación se presentan cuatro de las distribuciones más comunes que cumplen con dichas características.

#### **Distribución Binomial**

La distribución Binomial ocurre de manera natural al momento de analizar riesgos homogéneos, es decir, imaginando una cartera de  $n$  riesgos idénticos e independientes de los cuales nos interesa saber si se materializan o no.

Esta distribución es de gran importancia en la estadística gracias a sus diversas aplicaciones ya que en muchas áreas de estudio existen situaciones en las que se requiere analizar una muestra de observaciones en las que sólo existen dos resultados fácilmente clasificados en éxito y fracaso.

Un ejemplo de lo anterior sería una situación donde se tenga que estimar la probabilidad de que una muestra de la población elija entre dos opciones A y B como elecciones entre dos candidatos. En finanzas podríamos analizar instrumentos equivalentes, en seguros una cartera de  $n$  individuos con las mismas características de

mortalidad.<sup>1</sup> En todos los casos la probabilidad de que un elemento experimente un fallo está determinada por una distribución *Bernoulli*( $p$ ) con probabilidad  $1 - p$  en 0 y  $p$  en 1.

Es decir que para la cartera de  $n$  riesgos idénticos e independientes por analizar, la probabilidad de que existan  $k$  realizaciones del riesgo podría verse como observar  $n$  ensayos Bernoulli y que  $k$  de ellos resulten exitosos. Sin embargo, la probabilidad se verá afectada por el hecho de que existe más de un arreglo de la muestra en el que ocurren  $k$  éxitos dentro de los  $n$  ensayos.

Por tanto, la probabilidad de que ocurran exactamente  $k$  de los  $n$  riesgos está determinada por la variable aleatoria (v.a.)  $N$  con función de densidad:

$$P[N = k] = \binom{n}{k} p^k (1 - p)^{n-k}$$

Por lo que los valores de media y varianza para  $N \sim \text{Bin}(n, p)$  son:

$$\mathbb{E}[N] = np \quad \text{y} \quad \text{Var}[N] = np(1 - p).$$

Esta distribución tiene dos características importantes que apoyan al modelado de datos discretos. La primera es que, en términos nominales, la varianza es menor que la media por lo que si la media y varianza muestral se comportan de la misma manera, tendríamos un indicio de que la distribución que le ajusta de manera más significativa podría ser una Binomial.

La segunda característica es que tiene soporte finito, es decir que existe un entero  $M$  a partir del cual la distribución ya no presenta probabilidades positivas. Eso es particularmente útil para riesgos en los que no tiene sentido asignar probabilidad positiva a enteros mayores que un valor dado.

Por lo anterior, una de las aplicaciones más comunes de la distribución Binomial son los estudios médicos ya que en ellos generalmente se intenta medir la fatalidad de alguna condición médica. Por ejemplo, en el análisis de una muestra de pacientes que sufrieron un ataque cardíaco, el uso de esta distribución permitirá estimar la probabilidad de supervivencia.

---

<sup>1</sup>Siendo que el término *población*, puede ser entendido como una selección aleatoria de un conjunto más grande de elementos o la totalidad de ellos. En caso de que sea una muestra aleatoria de un total se debería analizar la representabilidad del total mediante el conjunto seleccionado, sin embargo es un tema extenso y no será desarrollado en este trabajo.

### Distribución Geométrica

Considérese una sucesión infinita de variables Bernoulli con probabilidad de éxito  $p$  y sea  $N$  v.a. que indique el primer ensayo en el que se obtiene un éxito. Entonces, la probabilidad de que el primer éxito se presente en el  $k$ -ésimo ensayo estará dada por la probabilidad de que en los primeros  $k - 1$  ensayos se obtengan fallos, cada uno con probabilidad de ocurrencia  $1 - p$ .<sup>2</sup>

El hecho de que exactamente los primeros  $k - 1$  ensayos sean fallos consecutivos implica que la probabilidad no es afectada algún factor combinatorio como en la distribución Binomial, por lo que la variable  $N$  tendrá función de densidad  $f_N(k) = (1 - p)^{k-1}p$  con  $k = 1, 2, 3, ..$  y se dice que se distribuye geoméricamente.

De manera similar la distribución de la variable  $Y = N - 1$  indica la probabilidad de que se realicen  $k$  ensayos fallidos antes del primer éxito; su función de densidad es  $f_Y(k) = (1 - p)^k p$  para  $k = 0, 1, 2, ..$

Para ambas variables, la esperanza y varianza están dadas por,

$$\mathbb{E}[N] = \frac{1}{p}, \quad \mathbb{E}[Y] = \frac{1-p}{p}, \quad \text{Var}[N] = \text{Var}[Y] = \frac{1-p}{p^2}.$$

La distribución geométrica posee la propiedad de pérdida de memoria, por lo que si describe el número de reclamaciones hechas, esta propiedad podría interpretarse como sigue:<sup>3</sup> “Dado a que existen  $m$  reclamaciones para un riesgo, la distribución de probabilidad de las reclamaciones que excedan  $m$  no dependen del valor de  $m$ .”

Es decir, descrito en términos de la probabilidad condicional, la pérdida de memoria se define como:

$$P[N > k + l | N > l] = P[N > k].$$

Otra herramienta de análisis que nos brinda esta distribución es que da un referente para determinar qué distribuciones discretas tienen colas pesadas o ligeras de la misma forma que lo hace la distribución exponencial con las distribuciones continuas.

Uno de los campos más importantes donde se encuentra esta distribución en la teoría de filas o colas.<sup>4</sup> Por ejemplo, en el caso de que exista una fila de espera, generalmente se supone que la probabilidad de que un *cliente*<sup>5</sup> llegue a la fila es  $p$  y que

<sup>2</sup>Charles M. Grinstead, J. Laurie Snell. (2006). Introduction to Probability. En Distributions and Densities(183-205). Estados Unidos: American Mathematical Society.

<sup>3</sup>Stuart A. Klugman, Harry H. Panjer, Gordon E. Willmot. (2004). Loss Models: From Data to Decisions. New Jersey, Estados Unidos: John Wiley & Sons.

<sup>4</sup>Del inglés “Queue theory”.

<sup>5</sup>Se hace uso de la palabra cliente ya que el uso más común de la teoría de filas es dentro del modelo de atención a clientes en una o varias cajas, sin embargo podría utilizarse prácticamente para cualquier fila *First In First Out*.

la probabilidad de que ningún cliente se incorpore a ella es  $1 - p$ .

Claramente la medición del tiempo no será de manera continua sino por bloques, es decir que por ejemplo en el caso de un banco puede determinarse como unidad de tiempo cada minuto del día, mientras que en un centro mecánico la unidad de tiempo podrá estar determinada por una hora o algún otro lapso.

### Distribución Binomial Negativa

Para desarrollar la distribución, consideremos una v.a.  $N$  que denota el número de veces que se realiza un experimento Bernoulli con probabilidad de éxito  $p$  hasta que se obtiene el  $k$ -ésimo éxito con  $k > 0$ .

Por lo tanto, la probabilidad de que  $N = n$  está dada por la probabilidad que en los primeros  $n - 1$  ensayos se tengan  $k - 1$  éxitos y  $n - k$  fallos y en el ensayo  $n$  se tenga el  $k$ -ésimo éxito,<sup>6</sup> de nuevo teniendo en mente las posibles formas en las que se pueden acomodar los éxitos y fracasos en el arreglo de ensayos, i.e.

$$P[N = n] = \binom{n-1}{k-1} p^k (1-p)^{n-k} \quad n \in \{k, k+1, \dots\}$$

Dicha probabilidad puede ser re-expresada como:

$$P[N = n] = \binom{n+k-1}{k-1} p^k (1-p)^n \quad n \in \{0, 1, \dots\}$$

Con lo cual, la media y varianza están dados por:

$$\mathbb{E}[N] = \frac{kp}{1-p} \quad y \quad Var[N] = \frac{kp}{(1-p)^2}.$$

Existe una tercera parametrización de la función de probabilidad en la que se considera la probabilidad de éxito de los ensayos Bernoulli como  $p = \frac{1}{1+\beta}$  y la cantidad de éxitos como  $r$ , con lo que para la variable  $N$  con distribución Binomial Negativa de parámetros  $r$  y  $\beta$ .

Dado que esta distribución depende de dos parámetros, da espacio a mayor flexibilidad en su forma lo que extiende su uso, por ejemplo, en el análisis de propagación de unidades. Esto es especialmente útil para la medicina y la ecología en el estudio del contagio de virus o parásitos dentro de un ambiente, ya que permite estimar la tasa de incidencia o disipación del elemento en propagación.

---

<sup>6</sup>Prasanna Sahoo.(2013). Some Special Discrete Distributions. En Probability and Mathematical Statistics(107-129). Estados Unidos: University of Louisville.

## Distribución Poisson

La distribución Poisson, toma el nombre del físico y matemático de origen francés Simeón-Denis Poisson<sup>7</sup> quien en 1838 publicaría en París su obra "*Recherches sur la probabilité des jugements en matières criminelles et matière civile*". Obra en la cual presentaría por primera vez esta distribución como una derivación de la distribución Binomial.

Para ver dicha relación, imaginemos un intervalo de tiempo continuo en el cual se marcan ocurrencias de cierto evento. Este proceso puede modelarse mediante una distribución binomial en tanto el tiempo sea dividido en  $n$  intervalos con la misma longitud tales que, dentro de cada uno de ellos pueda presentarse a lo más una única ocurrencia. Bajo estas consideraciones, el proceso podría verse como una sucesión de ensayos Bernoulli donde el éxito indica la ocurrencia del evento en ese intervalo.

Ya que se describe la ocurrencia de eventos en el tiempo, se puede tomar una tasa de ocurrencia, digamos  $\lambda$  tal que el número de ocurrencias en el intervalo de longitud  $t$  esté dado por  $\lambda t$  por lo que, en promedio se podrán esperar  $np = \lambda t$  eventos.

Así, si utilizamos la información de la distribución binomial que se presentó anteriormente y si consideramos la longitud de los sub-intervalos como 1 para la v.a.  $X$  que modela el proceso, se obtiene:

$$P[X = 0] = (1 - p)^n = \left(1 - \frac{\lambda}{n}\right)^n.$$

Se puede apreciar que si  $n \rightarrow \infty$ , entonces la probabilidad tiende a  $e^{-\lambda}$ .

Por otra parte, se tiene que

$$\frac{P[X = k]}{P[X = k - 1]} = \frac{\lambda - (k - 1)p}{kp}.$$

Para valores grandes de  $n$  será  $\frac{\lambda}{k}$ , pues  $p \rightarrow 0$ .

Por lo tanto, para  $k \in \{0, 1, 2, \dots\}$ , la función de densidad para una variable aleatoria  $N \sim \text{Poisson}(\lambda)$  es:

$$f_N(k) = \frac{e^{-\lambda} \lambda^k}{k!},$$

con valores de media y varianza:

$$\mathbb{E}[N] = \lambda \quad y \quad \text{Var}[N] = \lambda.$$

---

<sup>7</sup>Simeón-Denis Poisson, consulta 07/07/2018, fuente: <https://www.britannica.com/biography/Simeon-Denis-Poisson>

Que es una característica muy peculiar de esta distribución ya que entonces  $\mathbb{E}[N] = \text{Var}[N]$ , lo cual es muy útil para identificar la posible distribución de datos con media y varianza similares.

## 1.2. Distribuciones continuas

Ya se han presentado distribuciones para modelar variables de frecuencia por lo que ahora se expondrán distribuciones útiles para el modelado de severidad. Por la naturaleza de los riesgos a tratar, los montos de pérdidas o la cantidad de minutos de retraso en un vuelo son claramente aleatorios y no negativos, por ello se utilizan distribuciones que consideran lo siguiente:

- La variable  $X$  asociada a la severidad del riesgo es continua no negativa i.e.

$$X : [0, \infty) \rightarrow [0, \infty). \quad (1.1)$$

- La distribución es generalmente sesgada positiva i.e.

$$\frac{\mathbb{E}[(S - \mathbb{E}[S])^3]}{\text{Var}(S)^{3/2}} > 0. \quad (1.2)$$

Teniendo en cuenta estas características, es necesario conocer algunas de las familias paramétricas más adecuadas para el modelado de la severidad de riesgos. Con este fin, se tratarán a continuación algunas de distribuciones comúnmente utilizadas y que además fueron analizadas para el ajuste de los datos que se presentará en el cuarto capítulo de esta tesis.

### Distribución Exponencial

Sea  $X$  una v.a. con soporte sobre los reales no negativos, se dice que  $X$  se distribuye exponencialmente si su función de densidad está dada por:

$$f_X(x) = \theta e^{-\theta x}, \quad \text{con } \theta \text{ parámetro de escala.}$$

Esta distribución está estrechamente ligada a la distribución Poisson por su utilidad en la medición del tiempo de espera y algunas de sus características son:

$$\mathbb{E}[X] = \frac{1}{\theta}, \quad \text{Var}[X] = \frac{1}{\theta^2}, \quad m_X(t) = \left( \frac{\theta}{\theta - t} \right),$$

donde  $m_X(t)$  es la función generadora de momentos (f.g.m.).

Al igual que la distribución geométrica, la distribución exponencial posee la propiedad de pérdida de memoria, la diferencia esencial está en que, al ser continua, esta distribución brindará un mejor enfoque en sus aplicaciones en cálculo de tiempos.



Ejemplo de lo anterior son algunos usos dentro de la Investigación de Operaciones como el cálculo del tiempo medio entre fallos<sup>8</sup> usualmente utilizado en teoría de filas y controles de calidad.

La distribución exponencial funciona como referente para ver si la distribución de una variable aleatoria  $Y : [0, \infty) \rightarrow [0, \infty)$  tiene colas pesadas o ligeras, es decir, que si  $f_Y(y|\Omega)$  es la función de densidad de la v.a  $Y$  dados los parámetros  $\Omega = \{\omega_1, \dots, \omega_n\}$  y  $f_X(x|\theta)$  es la función de densidad para alguna v.a.  $X$  tal que  $X \sim Exp(\theta)$ , definimos la función  $G(X, Y)$  como:

$$G(X, Y) := \lim_{x \rightarrow \infty} \frac{f_Y(x|\Omega)}{f_X(x|\theta)}$$

Obteniendo un valor constante que, si es menor o igual a uno bajo ciertas circunstancias, indica que sus colas son al menos tan ligeras como la de la distribución exponencial, i.e.

$$G(X, Y) = \begin{cases} c \leq 1 & \text{colas ligeras.} \\ c > 1 & \text{colas pesadas.} \end{cases}$$

Algo importante de hacer notar es que para algunas distribuciones el valor del límite dependerá de  $\Omega$ , lo cual también brinda información sobre la distribución analizada como se mostrará más adelante en este capítulo.

Con la función anterior podemos determinar en qué condiciones la distribución de la v.a.  $Y$  tiene colas de mayor o menor peso, pero ¿qué significa esto? Y ¿en qué afecta a la modelación del riesgo? Las colas de las distribuciones con sesgo positivo nos muestran eventos que difieren en gran medida del valor de la media y que, en palabras más asociadas al riesgo analizado en esta tesis, quiere decir que son los casos en los que el tiempo de incumplimiento fue muy grande.

Por lo tanto, que un riesgo “sea” de colas pesadas quiere decir ya sea, que una proporción relativamente grande de las observaciones tienen valores grandes, o que existen observaciones de eventos con valores tan grandes que aunque no sean muchos cargan la probabilidad a la derecha.

Como observación adicional es oportuno comentar que la función  $G(X, Y)$  es únicamente un apoyo para el análisis de distribuciones y no deberá tomarse como regla única y absoluta.

---

<sup>8</sup>Del inglés “Mean Time Between Faliures”(MTBF).

### Distribución Gamma

Una v.a.  $X$  tiene distribución Gamma si su función de densidad es de la forma:

$$f_X(x) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} \quad \alpha \geq 0, \theta \geq 0.$$

Donde  $\alpha$  es un parámetro de forma,  $\theta$  es parámetro de escala y la función Gamma está dada por:

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \quad \text{para } z \in \mathbb{C}.$$

Tal que para  $\alpha \in \mathbb{N}$ ,  $\Gamma(\alpha) = (\alpha - 1)!$

Algunas de las características de la distribución Gamma son:

$$\mathbb{E}[X] = \frac{\alpha}{\theta}, \quad \text{Var}[X] = \frac{\alpha}{\theta^2}, \quad m_X(t) = \left( \frac{\theta}{\theta - t} \right)^\alpha.$$

Si se tienen  $X_1, X_2, \dots, X_n \sim \text{Exp}(\theta)$  Variables aleatorias independientes idénticamente distribuidas (v.a.i.i.d.), entonces  $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \theta)$  con función de distribución  $f_X(x) = \frac{\theta^n}{n!} x^{n-1} e^{-\theta x}$ , por lo que la distribución exponencial es el caso particular de la Gamma cuando  $\alpha = 1$ .

### Distribución Gamma Inversa

Sea  $X \sim \text{Gamma}(\alpha, \theta)$  y  $Y$  tal que  $Y = 1/X$ , entonces la distribución de  $Y$  está dada por:<sup>9</sup>

$$\begin{aligned} f_Y(y) &= f_X(1/y) \left[ \frac{d}{dy} y^{-1} \right] \\ &= \frac{1}{\theta^\alpha \Gamma(\alpha)} \left( \frac{1}{y} \right)^{\alpha-1} e^{-x/\theta} \left( \frac{1}{y} \right)^2 \\ &= \frac{1}{\theta^\alpha \Gamma(\alpha)} \left( \frac{1}{y} \right)^{\alpha+1} e^{-x/\theta} \\ &= \frac{y^{-\alpha+1}}{\theta^\alpha \Gamma(\alpha)} e^{-x/\theta}. \end{aligned}$$

Para el cálculo del  $n$ -ésimo momento de la distribución debe ser que  $\alpha > n$  por lo que la esperanza existe para  $\alpha > 1$  y es  $\mathbb{E}[X] = \frac{1}{(\alpha - 1)\theta}$  y para  $\alpha > 2$ ,

$$\text{Var}[X] = \frac{1}{(\alpha - 1)^2(\alpha - 2)\theta}.$$

<sup>9</sup>John D. Cook. (2008). Inverse Gamma Distribution., consulta 20/01/2019, fuente: [https://www.johndcook.com/inverse\\_gamma.pdf](https://www.johndcook.com/inverse_gamma.pdf)

Como se puede ver, la Gamma inversa deriva de la distribución Gamma y sus principales usos se encuentran dentro de la estadística bayesiana.

### Distribución Pareto

La distribución Pareto además de ser sesgada positiva, tiene cola pesada, es decir modela mejor los riesgos que generen pérdidas o incumplimientos mayores con más frecuencia. Esta distribución particular surge de considerar que la función de supervivencia de una v.a. tiene la forma:

$$S_X(x) = P[X > x] = \left(\frac{\theta}{x}\right)^\alpha \quad \text{para } \alpha > 0, \theta > 0.$$

De lo cual se obtiene que la función de densidad es de la forma:

$$f_X(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}},$$

Donde  $x > \theta$  y  $\alpha$  y  $\theta$  son parámetros de forma y escala respectivamente. Algunas de sus características son:

- $\mathbb{E}[X^r] = \frac{\alpha\theta^r}{(\alpha - r)}$  con  $\alpha > r, r = 1, 2, \dots$
- $\mathbb{E}[X] = \frac{\alpha\theta}{\alpha - 1}$  para  $\alpha > 1$ .
- $Var[X] = \frac{\alpha\theta^2}{(\alpha - 2)(\alpha - 1)^2}$  para  $\alpha > 2$ .

Esta distribución, comúnmente conocida como Pareto de primer tipo, tiene soporte únicamente en el intervalo  $[\theta, \infty)$ , por lo que no permitiría un correcto ajuste a riesgos que, dentro de su naturaleza, permitan valores de severidad entre 0 y  $\theta$ . Para enfrentar ese problema existe una segunda versión de la distribución, la cual es simplemente la traslación de la Pareto de primer tipo al origen, es decir, la v.a.  $Y = X - \theta$  se distribuye Pareto segundo tipo y su función de densidad es:

$$f_Y(x) = \frac{\alpha\theta^\alpha}{(x + \theta)^{\alpha+1}} \quad \text{para } x \geq 0, \alpha > 0, \theta > 0.$$

Además, esta distribución trasladada, conserva los mismos valores de esperanza y de varianza que la de primer tipo.

### Distribución Weibull

La distribución Weibull nombrada en honor al matemático sueco Walossi Weibull, está ampliamente relacionada con otras distribuciones tales como la distribución Generalizada de Valores Máximos y las distribuciones Gamma y Normal entre otras. Aquí se presenta la distribución como generalización de la exponencial mediante la

agregación de un parámetro de forma  $\alpha$ .<sup>10</sup>

Sea  $X \sim Exp(\theta = 1)$  con  $f(x) = e^{-x}$ , entonces la función de distribución de la v.a.  $Y = \left(\frac{X}{\theta}\right)^\alpha$  estará dada por:

$$\begin{aligned} f_Y(y) &= f_X((y/\theta)^\alpha) \left[ \frac{d}{dy}(y/\theta)^\alpha \right] \\ &= \frac{\alpha}{\theta} \left(\frac{y}{\theta}\right)^{\alpha-1} \exp\{-(y/\theta)^\alpha\}. \end{aligned}$$

Así, algunas propiedades de la distribución  $Weibull(\alpha, \theta)$  son:

- $\mathbb{E}[X] = \theta \Gamma\left(1 + \frac{1}{\alpha}\right)$ ,
- $Var[X] = \theta^2 \left[ \Gamma\left(2 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right]$ .

Una de las aplicaciones más comunes de esta distribución, es el modelado del tiempo de fallo en el que la tasa a la cual ocurren, es proporcional a un factor multiplicativo sobre el tiempo, por lo que, si en la distribución Weibull se fija el parámetro  $\alpha = 1$ , entonces se obtiene de nuevo la distribución  $Exp(\theta)$  en la cual la tasa de fallo es constante.

Si el parámetro es menor a uno ( $\alpha < 1$ ), entonces la tasa de fallo decrece conforme el tiempo avanza y por el contrario si ( $\alpha > 1$ ) entonces se le conoce como proceso de envejecimiento.<sup>11</sup>

### Distribución Beta Generalizada

La función de distribución Beta generalizada del segundo tipo fue ampliamente estudiada y aplicada en la segunda mitad del siglo XX y es la generalización más amplia de las distribuciones tipo Beta pues tiene cuatro parámetros;  $a, p$  y  $q$  son parámetros de forma y  $b$  de escala.

La densidad de una v.a.  $X \sim BG2(a, b, p, q)$  es:

$$f_X(x) = \frac{ax^{ap-1}}{b^{ap} \mathbf{B}(p, q) [1 + (x/b)^a]^{p+q}} \quad \text{para } x, a, b, p, q > 0.$$

Siendo la función Beta:  $\mathbf{B}(r, s) = \int_0^1 t^{r-1} (1-t)^{s-1} dt$  para  $r, s \in \mathbb{R}^+$ .

<sup>10</sup>Horst Rinne. (2008), Related distributions from: The Weibull Distribution, A Handbook. CRC Press

<sup>11</sup>Horst Rinne. (2011). A study of Weibull shape parameter: Properties and significance. Reino Unido: Elsevier.

Los tres parámetros de forma y el de escala permiten gran flexibilidad a esta distribución, además para casos particulares de cada uno de ellos, se derivan distribuciones con nombres y aplicaciones particulares como las dos presentadas a continuación.

### Distribución Burr

La distribución Burr fue presentada por primera vez por el matemático estadounidense Irving Wingate Burr en un texto sobre ecuaciones diferenciales en 1942. Esta distribución ha sido redescubierta en distintos momentos de la historia en ámbitos diferentes, prueba de ello, es que esta distribución también es comúnmente conocida como la distribución Singh–Maddala, ya que estos matemáticos ahondaron en sus características y aplicaciones entre 1975 y 1976.<sup>12</sup>

La distribución es el caso particular en la que el parámetro de forma  $p$  de la distribución  $BG2$  es igual a uno por lo que su densidad está dada por:

$$f_X(x) = \frac{ax^{a-1}}{b^a[1 + (x/b)^a]^{1+q}} \quad \text{para } x, a, b, q > 0.$$

Algunas de sus características son:

- $\mathbb{E}[X^r] = \frac{b^k \mathbf{B}(q - k/a, 1 + k/a)}{\mathbf{B}(1, q)}$
- $\mathbb{E}[X] = \frac{b \mathbf{B}(q - 1/a, 1 + 1/a)}{\mathbf{B}(1, q)} = \frac{b \Gamma(q - 1/a) \Gamma(1 + 1/a)}{\Gamma(q)}$
- $Var[X] = \frac{b^2 [\Gamma(q) \Gamma(1 + 2/a) \Gamma(q - 2/a) - \Gamma^2(1 + 1/a) \Gamma^2(q - 1/a)]}{\Gamma^2(q)}$

### Distribución Log-Logística

Existen diversos nombres para esta distribución dependiendo del área de estudio desde la cual sea estudiada, por ejemplo, en ámbitos económicos como el estudio de la distribución del ingreso en una sociedad se le conoce como la distribución Fisk. En el ámbito estadístico-matemático su nombre más difundido es Log-logística ya que si la v.a.  $X \sim Llogis(a, b)$  entonces  $Y = \log(X) \sim Logstica(a, \log(b))$ .<sup>13</sup>

Así mismo, la distribución Log-logística es un caso particular de la distribución Singh–Maddala, y por ende de la Beta generalizada, en la que el parámetro de forma  $q$  es uno.

<sup>12</sup>Christian Kleiber, Samuel Kotz. (2003). Beta-Type Size Distributions. En Statistical Size Distributions in Economics and Actuarial Sciences (183-230). Estados Unidos: John Wiley and Sons, Inc.

<sup>13</sup>Christian Kleiber, Samuel Kotz. (2003). Beta-Type Size Distributions. En Statistical Size Distributions in Economics and Actuarial Sciences (183-230). Estados Unidos: John Wiley and Sons, Inc.

La *f.d.p* de una variable con distribución Log-logística es:

$$f_X(x) = \frac{ax^{a-1}}{b^a[1 + (x/b)^a]^2} \quad \text{para } x, a, b > 0.$$

Algunas de sus características son:

- $\mathbb{E}[X^r] = b^k \mathbf{B}(1 - k/a, 1 + k/a)$  con  $a > k, r = 1, 2, \dots$
- $\mathbb{E}[X] = \frac{b\pi}{a \sin(\pi/a)}$  para  $a > 1$ .
- $Var[X] = b^2 \left( \frac{2a}{\sin(2\pi/a)} - \frac{a^2}{\sin^2(\pi/a)} \right)$  para  $a > 2$ .

### 1.3. Ajuste de distribuciones

Está claro que el proceso de modelado es únicamente una estimación de la realidad, por lo que al modelar un riesgo, no se trabaja con el riesgo real sino con una aproximación matemática de este. Así pues, el objetivo de un actuario es generar representaciones que brinden un ajuste suficientemente bueno de los datos para finalmente elegir aquella que, dentro de las limitaciones naturales existentes, permita dar la descripción más precisa de la realidad; sin embargo para poder decidir qué modelo es el “*mejor*” de entre todos existen criterios de ordenamiento.

Una forma práctica pero imprecisa de realizar el ajuste es comparando de manera gráfica, qué tan similares son las funciones de densidad o distribución del modelo y de los datos. Es obvio que este método es altamente subjetivo ya que depende de qué aprecie más el actuario; en el ejemplo de que dos actuarios realicen el análisis del mismo conjunto de datos, puede que uno elija gráficamente un modelo A pues aparenta ajustar mejor la cola de los datos, mientras que el otro elija uno B pues aunque las gráficas no son tan similares, la distribución es más fácil de trabajar.

A pesar de la practicidad del enfoque anterior, es bueno que exista mayor rigor matemático en la selección de los modelos. La forma más utilizada en estadística para resolver esa inquietud es el uso de las pruebas de bondad de ajuste.

En esencia una prueba de bondad de ajuste es una prueba de hipótesis en la cual se prueba una distribución específica y no así los valores de parámetros de una distribución, es decir:

$H_0$ : Los datos siguen una distribución  $F_0$ .

$H_1$ : Los datos siguen alguna distribución  $F$  diferente a  $F_0$ .

Lo cual se puede expresar como:

$$H_0 : F(x) = F_0(x) \quad \forall x \quad v.s. \quad H_1 : F(x) \neq F_0(x) \quad \text{para algún } x.$$

Donde  $F_0$  puede ser totalmente conocida, parcialmente desconocida o totalmente desconocida; es decir, dada la forma funcional propuesta puede conocerse el valor de todos los parámetros, de algunos de ellos o de ninguno.

Es propio mencionar que en caso de rechazar la hipótesis nula, únicamente se estará evidenciando que esa distribución particular no ajusta los datos y no se falla a favor de alguna distribución alterna.

Para lo anterior existen dos medidas principales: el nivel de significancia  $\alpha$  y el p-valor (en inglés *p-value*).<sup>14</sup>

El nivel de significancia  $\alpha$  de una prueba de hipótesis es la probabilidad de rechazar la hipótesis nula dado que es cierta. Si la hipótesis  $H_0$  puede ser verdadera en más de una manera, por ejemplo en pruebas con hipótesis compuestas, el nivel de significancia será el máximo de las probabilidades, por lo que:

$$\alpha = P[\text{Rechazar } H_0 | H_0].$$

Por otra parte, el p-valor es la probabilidad de obtener valores más extremos que los observados suponiendo cierta la hipótesis nula:

$$p - \text{value} = P[X \text{ es más extremo que } x | H_0].$$

En general, el valor de  $\alpha$  se da de antemano a las pruebas, queriendo que sea un valor pequeño, ya que si es grande la decisión de aceptar o rechazar  $H_0$  no es convincente pues se estaría dando lugar a una mayor probabilidad de tomar la decisión incorrecta.

A continuación se presentan dos de las pruebas de bondad de ajuste no paramétricas más utilizadas en la estadística que permiten probar si una muestra proviene de distribuciones continuas unidimensionales.

### Kolmogorov-Smirnov

La prueba Kolmogorov-Smirnov fue presentada por los matemáticos Andrey Kolmogorov y Nikolai Smirnov nacidos en la Unión Soviética, actual Rusia. Esta prueba es utilizada en dos contextos, uno es como prueba de bondad de ajuste para comparar una muestra con una distribución continua y el segundo es como prueba de comparabilidad entre las distribuciones de dos muestras para determinar si ambas provienen de la misma distribución teórica.<sup>15</sup>

La estadística de prueba está dada por:

$$D = \sup_x |F_n(x) - F_0(x)|.$$

<sup>14</sup>George Casella, Roger L. Berger. (2002). Hypothesis Testing. En Statistical Inference(423). California, Estados Unidos: Duxbury Advanced Series.

<sup>15</sup>En inglés “One-Sample K-S test” y “Two-Sample K-S test” respectivamente.

Donde  $F_n(x)$  es la distribución empírica de los datos y  $F_0(x)$  la distribución teórica.

El cálculo de la estadística  $D$  para una muestra  $X$  se realiza mediante las estadísticas auxiliares  $D^+$  y  $D^-$ , de la siguiente manera:

$$D^+ = \max_{1 \leq x \leq n} \left\{ \frac{1}{n} - F_0(x_i) \right\},$$

$$D^- = \max_{1 \leq x \leq n} \left\{ F_0(x_i) - \frac{(i-1)}{n} \right\},$$

$$D = \max\{D^+, D^-\}.$$

Donde  $x_i$  es el  $i$ -ésimo elemento en la muestra de datos ordenados.

Así pues, la estadística  $D$  representa la mayor diferencia entre la distribución empírica y la propuesta en la hipótesis nula.

### Anderson-Darling

En la prueba de Kolmogorov-Smirnov se toma como referencia la discrepancia entre  $F_n$  y  $F_0$ , sin embargo la intuición nos indica que aún es posible mejorarla, así pues, los estadistas americanos Theodore Wilbur Anderson y Donald A. Darling presentaron en 1952 la prueba que llevaría su nombre.

La familia de pruebas Cramer-Von Mises es el conjunto de pruebas de bondad de ajuste que incorporan un sistema de ponderaciones a la diferencia cuadrática entre las distribuciones. Su estadística está dada por:<sup>16</sup>

$$W^* = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 \psi[F_0(x)] dF_0.$$

Donde la función  $\psi$  es el ponderador.

La estadística de prueba para Anderson-Darling es el caso particular en que se da mayor peso a las colas de la distribución con la función  $\psi(u) = \frac{1}{u(1-u)}$  por lo que la estadística es:<sup>17</sup>

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} dF_0(x).$$

<sup>16</sup>T. W. Anderson. (2010). The Anderson-Darling Statistic. En Anderson-Darling Tests of Goodness-of-Fit(8). California, Estados Unidos: Stanford University.

<sup>17</sup>Michael A. Stephens. (1979). The Anderson-Darling Statistic. California Estados Unidos: Stanford University.



### Pruebas de Razón de verosimilitud

En ocasiones no basta probar el nivel de significancia en el ajuste de una distribución, sino que se busca comparar de manera explícita dos modelos distintos. Ante esta necesidad surgen los test o pruebas basadas en la razón entre las funciones de verosimilitud de los modelos y que contrastan un modelo nulo y uno alternativo.

La Verosimilitud (del inglés *Likelihood*) es la probabilidad de observar los datos de una muestra en función de los parámetros. Así, para  $x_1, x_2, \dots, x_n$  observaciones de una variable aleatoria  $X$ , la función de verosimilitud está dada por:

$$L(\Theta|x) = \prod_{i=1}^n f(x_i, \Theta).$$

La razón de verosimilitud o *Likelihood ratio* es una función sobre una muestra y un espacio paramétrico que representa qué tanto es más factible que las observaciones de la muestra provengan de uno u otro modelo. Las pruebas de razón de verosimilitud, al igual que las pruebas de hipótesis utilizadas para la inferencia puntual, pueden ser tanto simples como compuestas, es decir, se pueden probar entre sí dos valores particulares o contrastar un elemento o conjunto contra su complemento. Para pruebas de hipótesis compuestas la razón de verosimilitud está definida como,

**Definición 1.3.1.** Sea  $x = \{x_1, x_2, \dots, x_n\}$  observaciones de una variable aleatoria  $X$ ,  $\Theta$  un espacio paramétrico y  $\Theta_0$  un subconjunto de dicho espacio, la razón de verosimilitudes está dada por

$$\Lambda(x) = \frac{\sup\{L(\theta|x); \theta \in \Theta_0\}}{\sup\{L(\theta|x); \theta \in \Theta\}} \quad (1.3)$$

Siguiendo esa definición  $\Lambda(x) \in (0, 1)$  y se acerca a cero si el modelo alternativo es mejor que el modelo nulo por lo que la regla de decisión lógica es rechazar  $H_0$  en favor de  $H_1$  si el valor de la razón está entre 0 y algún valor  $c$  propuesto como umbral de rechazo. Por lo tanto una prueba de razón de verosimilitud está definida como cualquier prueba que tenga una región de rechazo de la forma  $\{x : \Lambda(x) \leq c\}$ , donde  $c$  es cualquier número real entre 0 y 1.

# Capítulo 2

## Modelo de Mezclas y algoritmo EM

### 2.1. Mezclas de distribuciones

Como se vio anteriormente, dependiendo del nivel de significancia tomado como referencia para las pruebas, puede que el ajuste de una distribución a los datos no sea bueno y no represente de manera adecuada la información. Existen casos en los que no es posible modelar los datos mediante el ajuste de alguna de las densidades más conocidas como distribuciones con puntos de densidad o distribuciones multimodales.

Los modelos de mezclas ofrecen posibilidades más amplias que las distribuciones simples ya que, aunque la mezclas se componen de estas mismas distribuciones, su alta maleabilidad permite lograr un mejor ajuste a datos con formas menos convencionales.

Las mezclas de distribuciones son relativamente comunes ya que suelen presentarse incluso en situaciones de la vida cotidiana, sobre todo en riesgos o eventos que generen información no homogénea, por ejemplo si una empresa aseguradora incluye en una misma cartera la cobertura de riesgo de vida individual reclamaciones procedentes de una rama en Japón y de una en México, resulta intuitivo que los datos obtenidos de dicha cobertura, a pesar de proceder del mismo riesgo, pueden tener características un tanto diferentes.

**Definición 2.1.1.** Dadas  $X_1, X_2, \dots, X_k$  variables aleatorias independientes con funciones de densidad  $f_1, f_2, \dots, f_k$  y el conjunto de pesos  $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_k\}$  tales que:  $0 < \lambda_j < 1 \forall j \in \{1, \dots, k\}$  y  $\sum_{j=1}^k \lambda_j = 1$ . Entonces la f.d.p. de una mezcla de distribuciones con  $k$  componentes está dada por la combinación lineal convexa de las  $k$  densidades, i.e.:

$$f_M(x) = \lambda_1 f_1(x) + \lambda_2 f_2(x) + \dots + \lambda_k f_k(x),$$
$$f_M(x) = \sum_{j=1}^k \lambda_j f_j(x). \quad (2.1)$$

El hecho de que una mezcla esté dada por una combinación lineal convexa, implica que los momentos de la mezcla  $M$  están dados también por una combinación de los momentos de las variables  $X_j$  dados por:

$$\mathbb{E}[M^m] = \sum_{j=1}^k \lambda_j \mathbb{E}[X_j^m].$$

Por lo que la esperanza está determinada por:

$$\mathbb{E}[M] = \sum_{j=1}^k \lambda_j \mathbb{E}[X_j]. \quad (2.2)$$

Es importante mencionar que estas propiedades no dependen de qué distribuciones tengan las variables aleatorias, por lo que la mezcla puede estar compuesta por hasta  $k$  distribuciones donde cada una de las  $f'_j$ s sea una función de densidad única; sin embargo, en la práctica la mayoría de los casos se opta por ajustes por medio de modelos paramétricos, en los que las mezclas están compuestas por densidades  $f'_j$ s que pertenecen a la misma familia paramétrica pero con parámetro desconocido  $\theta_j$ .

Es decir, el modelo paramétrico de mezcla estaría dado por:<sup>1</sup>

$$f_M(x|\theta) = \sum_{j=1}^k \lambda_j f_j(x|\theta_j).$$

El vector de parámetros de la mezcla:  $\Theta = (\lambda_1, \dots, \lambda_k, \theta_1, \dots, \theta_k)$ .

Por sus características, las mezclas son más flexibles que una distribución única, sin embargo ganar mayor flexibilidad en el modelo implicará entrar en un proceso de *negociación* en el cual se intercambia la flexibilidad y capacidad de ajuste ganada mediante el uso de más densidades en el modelo por el incremento de la varianza derivado del cálculo de más parámetros.

## Identificabilidad

Antes de comenzar con los procesos de estimación paramétrica y no paramétrica de mezclas de distribuciones, se tiene que mencionar un detalle que puede generar problemas de interpretación: La identificabilidad del modelo.

Un modelo es identificable siempre que, “*dadas dos representaciones distintas de un modelo, se vean representadas dos estructuras muestrales distintas*”<sup>2</sup>, por lo tanto, si existen distintas representaciones de un modelo que sean equivalentes, se tiene que

<sup>1</sup>Carnegie Mellon University Department of Statistics. Guidelines for Advanced Data Analysis. En Mixture Models(30). Pittsburgh, Estados Unidos: Carnegie Mellon University.

<sup>2</sup>Carnegie Mellon University Department of Statistics. Guidelines for Advanced Data Analysis. En Mixture Models(4). Pittsburgh, Estados Unidos: Carnegie Mellon University.

cambiar la estructura de alguna de ellas o convenir una única representación fija.

Sabemos que una característica fundamental de las mezclas es su invariabilidad ante permutaciones de los índices de los componentes, es decir, es lo mismo si se estructura una mezcla de  $k$  componentes con la distribución  $f_j$  como  $n$ -ésima componente y la distribución  $f_i$  como la  $m$ -ésima a si se estructura con  $f_i$  como  $n$ -ésimo factor y  $f_j$  como el  $m$ -ésimo. Así,

$$\begin{aligned} f_{M'} &= f_1, \dots, f_{n-1}, f_j, f_{n+1}, \dots, f_{m-1}, f_i, f_{m+1} + \dots + f_k \\ f_{M''} &= f_1, \dots, f_{n-1}, f_i, f_{n+1}, \dots, f_{m-1}, f_j, f_{m+1} + \dots + f_k \\ &\Leftrightarrow f_{M'} = f_{M''} \end{aligned}$$

De manera general este fenómeno se conoce con el nombre de “*Label degeneracy*” o “*Label switching*” dependiendo de la fuente literaria de consulta.

### 2.1.1. Mezclas normales

Como sabemos, la distribución normal es una de las más conocidas y más utilizadas en la estadística ya que sus características la hacen fácil de emplear. Por eso mismo, uno de los modelos de mezclas más extendidos es el de  $k$  distribuciones normales. A continuación se presentan algunos de los resultados anteriores adaptados a estos modelos.

**Definición 2.1.2.** Dadas  $X_1, X_2, \dots, X_k$  variables aleatorias independientes tales que la v.a.  $j$ -ésima tiene distribución  $N(\mu_j, \sigma_j^2)$ . Entonces para los pesos  $\{\lambda_1, \lambda_2, \dots, \lambda_k\}$  la f.d.p. de una mezcla normal de  $k$  componentes está dada por:

$$f_M(x) = \sum_{j=1}^k \lambda_j N(\mu_j, \sigma_j^2). \quad (2.3)$$

Los modelos de mezclas de normales alcanzan gran flexibilidad incluso con pocas componentes al mismo tiempo que mantienen gran simplicidad en los resultados del modelo. Por ejemplo, la esperanza y varianza de la mezcla están dadas por:

- $\mathbb{E}[M] = \sum_{j=1}^k \lambda_j \mathbb{E}[X_j] = \sum_{j=1}^k \lambda_j \mu_j.$
- $Var[M] = \sum_{j=1}^k \lambda_j (\mu_j^2 + \sigma_j^2) - (\sum_{j=1}^k \lambda_j \mu_j)^2.$

#### **Demostración. Esperanza**

De la ecuación (2.2) se tiene que  $\mathbb{E}[M] = \sum_{j=1}^k \lambda_j \mathbb{E}[X_j].$

Por otro lado, para cada  $j \in \{1, 2, \dots, k\}$ ,  $\mathbb{E}[X_j] = \mu_j.$

Por lo tanto  $\mathbb{E}[M] = \sum_{j=1}^k \lambda_j \mu_j$  ■

**Demostración. Varianza**

La varianza puede expresarse como

$$\text{Var}[M] = \mathbb{E}[M^2] - \mathbb{E}^2[M].$$

Recordando que  $\mathbb{E}[M^m] = \sum_{j=1}^k \lambda_j \mathbb{E}[X_j^m]$  se tiene,

$$\begin{aligned} \text{Var}[M] &= \sum_{j=1}^k \lambda_j \mathbb{E}[X_j^2] - \mathbb{E}^2[M] \\ &= \sum_{j=1}^k \lambda_j \mathbb{E}[X_j^2] - \left( \sum_{j=1}^k \lambda_j \mu_j \right)^2 \\ &= \sum_{j=1}^k \lambda_j (\mu_j^2 + \sigma_j^2) - \left( \sum_{j=1}^k \lambda_j \mu_j \right)^2 \quad \blacksquare \end{aligned}$$

En este modelo es muy importante marcar de manera explícita la diferencia entre una mezcla y una suma de variables. Si se considera la suma  $X_1 + \dots + X_n$ , su distribución será de nuevo una distribución normal con parámetros  $N(\mu = \sum_{j=1}^n \mu_j, \sigma^2 = \sum_{j=1}^n \sigma_j^2)$ , sin embargo la distribución de la mezcla de normales no es normal ya que, si las medias de las componentes están alejadas entre sí, la distribución es multimodal que es claramente diferente a la unimodalidad y simetría características de la distribución normal.

## 2.2. Algoritmo EM

Existen diversos métodos para realizar estimaciones paramétricas como el método de Momentos y el de Mínimos Cuadrados pero también, que es buena idea utilizar el método de Máxima Verosimilitud siempre que sea posible. Pero ¿cómo se puede trasladar dicho método al ajuste de una mezcla de distribuciones? de ello trata el algoritmo que se presenta a continuación.

El algoritmo EM es un *proceso iterativo para el cálculo de estimaciones mediante la maximización de la función de verosimilitud cuando las observaciones pueden ser vistas como datos incompletos*.<sup>3</sup> Dicho algoritmo fue presentado en 1977 por los profesores Arthur P. Dempster, Nan Laird y Donald Rubin en el artículo “*Maximum Likelihood from Incomplete Data via the EM Algorithm*”.

El algoritmo de optimización, en el cual se basa el algoritmo EM, ha probado ser útil para diversas aplicaciones sobre todo en campos como *data clusterig, machine*

---

<sup>3</sup>Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). “Maximum Likelihood from Incomplete Data via the EM Algorithm”. *Journal of the Royal Statistical Society, Series B*

*learning, item response theory* y *big data* entre otros. En este trabajo se analizará como una herramienta de estimación del parámetro global  $\Theta = (\lambda_1, \dots, \lambda_k, \theta_1, \dots, \theta_k)$ .

Para desarrollar el algoritmo EM, se tiene que esclarecer primero el concepto de *datos incompletos*. Este término es utilizado por Dempster, Laird y Rubin para implicar la existencia de dos espacios muestrales, por ejemplo  $\mathcal{X}$  y  $\mathcal{Y}$  tales que los datos observados  $\mathbf{y}$  corresponden a  $\mathcal{Y}$ , mientras que los datos  $\mathbf{x}$  en  $\mathcal{X}$  sólo pueden ser observados indirectamente a través de un  $\mathbf{y}$ .

Introduzcamos el concepto de variables ocultas dentro del modelo, así podremos decir que:

$$P[X = x|\Theta] = \sum_z P[X = x|Z = z, \Theta]P[Z = z|\Theta]. \quad (2.4)$$

donde claramente  $Z$  funge como la variable oculta, de la cual no conocemos el valor.

Teniendo en cuenta la ecuación (2.4), podemos enfocar el algoritmo a la estimación de parámetros para una mezcla si pensamos en las constantes  $\lambda_1, \lambda_2, \dots, \lambda_K$  como las probabilidades de que la observación  $x_j$  pertenezca a la  $j$ -ésima distribución que compone la mezcla, entonces podremos pensar a  $\lambda$  como variable aleatoria con función de densidad  $g_\lambda(\lambda_j)$  y se cumple que  $\sum_\lambda g_\lambda(\lambda_j) = 1$  y  $0 < g(\lambda_j) \leq 1$ .

Por lo que la probabilidad que  $X = x_0$  depende tanto de la probabilidad que la observación  $x_0$  pertenezca a la  $j$ -ésima distribución como de que la distribución tome el valor  $x_0$ . Así para cualquier valor  $x$ :

$$P[X = x|\Theta] = \sum_j P[X = x|\lambda = \lambda_j, \Theta]P[\lambda = \lambda_j|\Theta]. \quad (2.5)$$

Por otra parte, se sabe que la verosimilitud es de la forma  $L(\Theta) = \prod_{i=1}^n f(x_i, \Theta)$ , por lo que, al aplicarle la función logaritmo para obtener la Log-Verosimilitud se obtiene:

$$l(\Theta) = \sum_{i=1}^n \ln f(x_i, \Theta).$$

Retomando la ecuación anterior y agregándola a la Log-Verosimilitud, se obtiene:

$$l(\Theta) = \sum_{i=1}^n \ln \sum_{k=1}^K \lambda_k f(x_i, \theta_k). \quad (2.6)$$

Derivando respecto a  $\theta_j$ :

$$\begin{aligned}
\frac{dl}{d\theta_j} &= \sum_{i=1}^n \frac{1}{\sum_{k=1}^K \lambda_k f(x_i, \theta_k)} \lambda_j \frac{df(x_i, \theta_k)}{d\theta_j} \\
&= \sum_{i=1}^n \frac{\lambda_j}{\sum_{k=1}^K \lambda_k f(x_i, \theta_k)} \frac{f(x_i, \theta_j)}{f(x_i, \theta_j)} \frac{df(x_i, \theta_k)}{d\theta_j} \\
&= \sum_{i=1}^n \frac{\lambda_j f(x_i, \theta_j)}{\sum_{k=1}^K \lambda_k f(x_i, \theta_k)} \frac{d \ln f(x_i, \theta_k)}{d\theta_j}.
\end{aligned}$$

En contraste, si se tomara la derivada de la log-verosimilitud de un modelo simple en el que no se incluya el efecto de las variables ocultas se obtendría:

$$\sum_{i=1}^n \frac{d \ln f(x_i, \theta_k)}{d\theta_j}.$$

Recordando que  $j \in \{1, \dots, K\}$  vemos de forma intuitiva que el proceso de maximización de la verosimilitud de un modelo de mezclas es comparable a la maximización ponderada de un modelo simple en donde los pesos están dados por

$$w_{ij} = \frac{\lambda_j f(x_i, \theta_j)}{\sum_{k=1}^K \lambda_k f(x_i, \theta_k)}.$$

Si bien esta forma de ver la verosimilitud de la mezcla ayuda a estructurar la maximización, también obvia el problema principal del proceso: los pesos dependen de los parámetros a estimar.

Para poder llegar a la solución del problema, se necesita comprender qué representan los pesos que se estructuraron. En el numerador se tiene  $\lambda_j f(x_i, \theta_j)$  que representa la probabilidad de que la observación provenga de la  $j$ -ésima distribución de la mezcla y que esa distribución en particular tome el valor  $x_i$ ; en el denominador está la probabilidad marginal  $P[X = x_i]$ .

Por lo tanto, los pesos estarán dados por:

$$w_{i,j} = P[\lambda = \lambda_j | X = x_i, \theta].$$

En pocas palabras, para estimar los parámetros de la mezcla mediante máxima verosimilitud habrá que estimar los pesos  $w_{i,j}$  pero para estimarlos se deben conocer los parámetros que en un principio se querían estimar, lo cual genera un problema circular de obtención de información. Para evitar eso se utiliza el proceso EM, pues el enfoque que aplica ataca este problema para romper el círculo vicioso de estimación paramétrica.

Supóngase que se tiene una asignación inicial del parámetro  $\Theta$  como  $\lambda_1, \dots, \lambda_K$  y  $\theta_1, \dots, \theta_n$ , con base en esos valores se puede calcular un valor para cada uno de los

pesos  $w_{i,j}$ . Una vez conociendo los pesos que se requieren, se puede realizar la maximización de la verosimilitud.

De esta primera “ejecución” de la maximización, se obtienen nuevos valores para el vector  $\Theta$  que, aunque no son los que maximizan la verosimilitud de la mezcla en un sentido final, sí son aquellos que generan la mayor verosimilitud posible dado los valores asignados. Lo que es más importante es que este nuevo vector  $\Theta'$  pueden utilizarse para calcular nuevos valores de los pesos  $w_{i,j}$  con base en los cuales se puede volver a realizar la maximización de la verosimilitud obteniendo nuevos valores del vector de parámetros.

Con lo anterior ya se puede vislumbrar el camino que sigue el algoritmo EM y es que uno llega a imaginar un proceso sucesivo de aproximación que debiera converger a los parámetros  $\hat{\lambda}_1, \dots, \hat{\lambda}_K$  y  $\hat{\theta}_1, \dots, \hat{\theta}_K$ .

Así el Algoritmo EM puede escribirse en los siguientes pasos:

1. Fijar los valores  $\lambda_1, \dots, \lambda_K, \theta_1, \dots, \theta_n$  del vector  $\Theta$ .<sup>4</sup>
2. Dados los parámetros fijos, calcular los pesos  $w_{ij} = \frac{\lambda_j f(x_i, \theta_j)}{\sum_{k=1}^K \lambda_k f(x_i, \theta_k)}$ .
3. Con los pesos obtenidos, maximizar la verosimilitud ponderada para obtener un nuevo vector de parámetros  $\Theta'$ .
4. Fijar los nuevos valores  $\Theta = \Theta'$  y regresar al paso (2).

El nombre del algoritmo proviene del inglés *Expectation-Maximization* y fue concebido así ya que el núcleo del algoritmo son el paso E, en el cual se calculan las probabilidades condicionales  $P[\lambda = \lambda_j | X = x_i, \theta]$  que es el cálculo de la esperanza de funciones indicadoras, y el paso M de maximización de la verosimilitud.

Ahora bien, hasta este momento desarrollamos el algoritmo EM enfocándolo al ajuste de mezclas y utilizando la notación propia para ello; sin embargo, el algoritmo EM es una herramienta general para la estimación de verosimilitud para modelos con variables ocultas. A continuación se desarrollará el algoritmo de manera más general, retomando la notación tal que  $Z$  denote la variable de información oculta y  $X$  la información explícita.

---

<sup>4</sup>Los valores iniciales del vector de parámetros son completamente subjetivos, sin embargo, existen ciertas recomendaciones lógicas para su asignación. Asignaciones que den peso cero a una o varias distribuciones son carentes de sentido práctico ya que los valores del vector  $\Lambda$  que generen el máximo de verosimilitud serán fraccionarios entre 0 y 1. Además, si se quisiera asignar pesos nulos sería más eficiente que desde un inicio se busque un ajuste únicamente con la cantidad de distribuciones deseadas.



Se desea encontrar el parámetro  $\Theta$  que maximice la verosimilitud, al igual que en desarrollo pasado, la maximización se apoya de la log-verosimilitud resultando,

$$L(\Theta) = \ln f(X|\Theta). \quad (2.7)$$

Además, ya se ha visto que el algoritmo es una maximización iterativa con lo que, para cada paso del proceso se estima un nuevo vector paramétrico que se acerca cada vez más a un máximo. Sea  $\Theta_n$  el vector obtenido de la  $n$ -ésima iteración del algoritmo y  $\Theta$  el vector que maximiza  $L$ .

$$L(\Theta) > L(\Theta_n).$$

Para probar que efectivamente las iteraciones tienden a un máximo se utiliza la siguiente ecuación.

$$L(\Theta) - L(\Theta_n) = \ln f(X|\Theta) - \ln f(X|\Theta_n). \quad (2.8)$$

Incorporando la variable de información oculta.

$$f(x|\Theta) = \sum_z f(X|Z, \theta) f(z|\theta). \quad (2.9)$$

Por lo que la ecuación (2.8) se puede reescribir como sigue,

$$L(\Theta) - L(\Theta_n) = \ln \sum_z f(X|z, \theta) f(z|\theta) - \ln f(X|\Theta_n). \quad (2.10)$$

Como  $Z$  es variable aleatoria, las probabilidades  $P[Z = z|X, \theta_n]$  son valores constantes y  $f(z|X, \theta_n) \geq 0$  para todos los valores  $z$  y  $\sum_z f(z|X, \theta_n) = 1$ . Esto es esencialmente lo que se hizo al introducir  $g(\cdot)$  como la función de probabilidad de los pesos  $\lambda$  de las mezclas. De la ecuación anterior se sigue,

$$\begin{aligned} L(\Theta) - L(\Theta_n) &= \ln \sum_z f(X|Z, \theta) f(z|\theta) - \ln f(X|\Theta_n) \\ &= \ln \sum_z f(X|Z, \theta) f(z|\theta) \left( \frac{f(z|X, \theta_n)}{f(z|X, \theta_n)} \right) - \ln f(X|\Theta_n) \\ &= \ln \sum_z f(z|X, \theta_n) \left( \frac{f(X|Z, \theta) f(z|\theta)}{f(z|X, \theta_n)} \right) - \ln f(X|\Theta_n). \end{aligned} \quad (2.11)$$

Dado el Teorema de Jensen y considerando que una función  $f$  es cóncava (estrictamente cóncava) si y sólo si  $-f$  es convexa (estrictamente convexa), tenemos que la Desigualdad de Jensen también es aplicable para funciones cóncavas pero cambiando la ‘dirección’ de las desigualdades; es decir,<sup>5</sup> para una función  $f$  una función *cóncava* tal que  $f''(x) \leq 0$  ( $\forall x \in \mathbb{R}$ ) y  $X$  variable aleatoria. Entonces:

$$f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)]. \quad (2.12)$$

<sup>5</sup>Sean Borman. (2004). The Expectation Maximization Algorithm A short tutorial. Utah, Estados Unidos: University of Utah, School of computing.

Aplicando la desigualdad de Jensen sobre la función logaritmo que es función cóncava, tenemos que:

$$\begin{aligned} L(\Theta) - L(\Theta_n) &= \ln \sum_z f(z|X, \theta_n) \left( \frac{f(X|Z, \theta) f(z|\theta)}{f(z|X, \theta_n)} \right) - \ln f(X|\Theta_n) \\ &\geq \sum_z f(z|X, \theta_n) \ln \left( \frac{f(X|Z, \theta) f(z|\theta)}{f(z|X, \theta_n)} \right) - \ln f(X|\Theta_n) \\ &= \sum_z f(z|X, \theta_n) \ln \left( \frac{f(X|Z, \theta) f(z|\theta)}{f(z|X, \theta_n) f(X|\Theta_n)} \right). \end{aligned}$$

Reduciendo la notación y pasando al lado derecho la verosimilitud de la  $n$ -ésima iteración se presenta la relación,

$$L(\Theta) \geq L(\Theta_n) + \Delta_n.$$

Lo cual proporciona un límite inferior para la Log-Verosimilitud, que siguiendo la notación utilizada por Sean Borman en su texto *“The Expectation Maximization Algorithm A short tutorial”*, se representará como  $L(\Theta_n) + \Delta_n = l(\theta|\theta_n)$ .

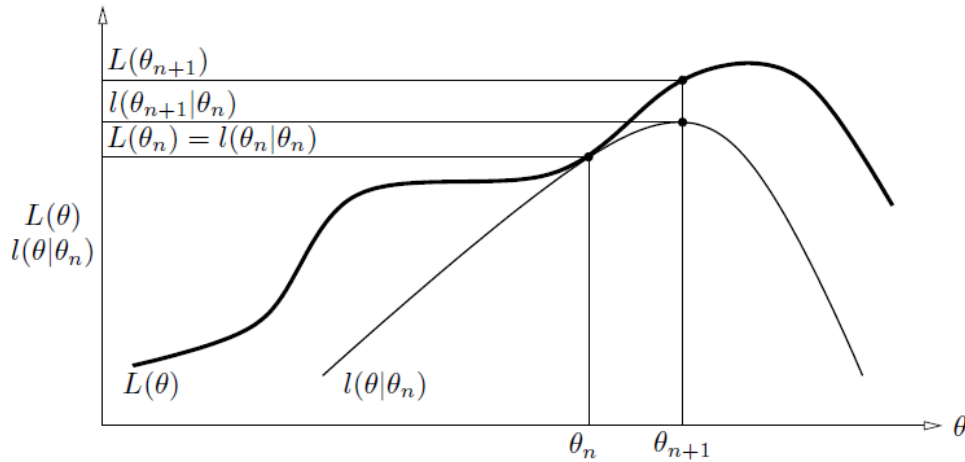


Figura 2.1: Representación gráfica de la  $n$ -ésima iteración

Donde  $\theta_{n+1}$  es el parámetro que maximiza  $l(\theta|\theta_n)$  y que se toma como valor base para la siguiente iteración del algoritmo. Este valor se obtiene como descrito a continuación.

$$\begin{aligned}
\theta_{n+1} &= \operatorname{argmax}_{\theta} \{l(\theta|\theta_n)\} \\
&= \operatorname{argmax}_{\theta} \left\{ L(\Theta_n) + \sum_z f(z|X, \theta_n) \ln \left( \frac{f(X|Z, \theta) f(z|\theta)}{f(z|X, \theta_n) f(X|\Theta_n)} \right) \right\} \\
&\text{Eliminando los valores constantes} \\
&= \operatorname{argmax}_{\theta} \left\{ \sum_z f(z|X, \theta_n) \ln (f(X|Z, \theta) f(z|\theta)) \right\} \\
&\text{Desarrollando las distribuciones condicionales,} \\
&= \operatorname{argmax}_{\theta} \left\{ \sum_z f(z|X, \theta_n) \ln \frac{f(X|Z, \theta) f(z|\theta)}{f(z|\theta) f(\theta)} \right\} \\
&= \operatorname{argmax}_{\theta} \left\{ \sum_z f(z|X, \theta_n) \ln (f(X, z|\theta)) \right\}.
\end{aligned}$$

Con lo cual se puede reescribir los pasos E y M del algoritmo como:

*Paso E:* Dado el parámetro  $\theta_n$ , calcular el valor de la esperanza condicional  $\sum_z f(z|X, \theta_n) \ln (f(X, z|\theta))$ .

*Paso M:* Calcular el parámetro  $\theta_{n+1} = \operatorname{argmax}_{\theta} \{ \sum_z f(z|X, \theta_n) \ln (f(X, z|\theta)) \}$ .

### Convergencia

Se desarrolló analíticamente el proceso de convergencia del algoritmo EM a un máximo de la verosimilitud, sin embargo, la convergencia presenta ciertos problemas más finos que son importantes a la hora de hacer un ajuste práctico.

Naturalmente, el hecho de que  $\theta_{n+1}$  sea el parámetro que maximice la diferencia  $\Delta_n$  provoca que para cada iteración la verosimilitud sea no decreciente. Sin embargo, en general la función de verosimilitud  $L$  puede tener formas no convencionales en las que se presenten diversos máximos locales.

Así pues, el proceso de maximización podría estar convergiendo a un máximo local, dependiendo de los valores sobre los que se inicien las iteraciones. Por lo tanto, la convergencia a un máximo global no está garantizada.<sup>6</sup>

Otro punto muy importante es que el algoritmo a pesar de tener fundamentos teóricos de convergencia, en la práctica los cálculos realizados en cada iteración se realizan mediante procesos numéricos, causando que los resultados obtenidos de la aplicación del algoritmo estén atados a las problemáticas de dichos procesos computacionales. Así en vez de que el algoritmo itere hasta conseguir estimar los parámetros que maximicen la verosimilitud, se realiza únicamente un número  $N$  de iteraciones

<sup>6</sup>Wu, C. F. Jeff. (1983). On the Convergence Properties of the EM Algorithm. En The Annals of Statistics. Vol 11 (95-103). Estados Unidos: Institute of Mathematical Statistics.

antes de terminar el proceso.

Es decir, el algoritmo para hasta que encuentra una cláusula de quiebre, que generalmente es hasta que la diferencia entre  $L(\theta_n)$  y  $L(\theta_{n+1})$  sea pequeña o hasta se cumpla un número grande de iteraciones. Lo anterior, como muchas otras situaciones en la estadística en la que se tiene que dar un parámetro de suficiencia, ya sea para pruebas, aceptaciones, estimaciones, entre otras, está rodeada de discusiones y subjetividad por parte de los estadísticos para intentar definir hasta qué punto la diferencia en valores de verosimilitud es lo suficientemente pequeña como para considerarse convergencia. Este punto será tratado de nuevo en el último capítulo.

## Capítulo 3

# El modelo agregado y sus aproximaciones

Modelar riesgos es parte fundamental de la labor del actuario y consiste de un proceso lógico que se realiza con la finalidad elegir una representación matemática de un suceso real.

Dicho proceso inicia con la selección de uno o varios posibles modelos que el observador considera pertinentes dado su criterio y conocimiento, posteriormente, calibrará los modelos con base en la información disponible y así estos modelos ya calibrados deberán ser probados para determinar qué tan adecuados son para describir los datos mediante el uso de diversas pruebas. Justamente para poder iniciar el proceso de modelado que describe esta tesis se necesita dar información que ponga en contexto modelo utilizado.

Como se mencionó al inicio de este trabajo, muchas de las herramientas actuariales que nacieron a causa de las necesidades del sector asegurador o fueron de resultados teóricos desarrollados en contextos diferentes y que se adaptaron al sector, así a lo largo del tiempo se desarrollaron distintos enfoques y formas de analizar los eventos y riesgos de interés para el sector asegurador; en particular son dos modelos los que cobraron mayor importancia y expandieron su uso tanto dentro del sector como fuera de él.

Uno de ellos es el *Modelo individual* que surgió de la búsqueda del monto de reclamación total de un portafolio de pólizas, ya que las compañías manejan una cantidad, digamos  $n$  de pólizas individuales que conforman la cartera de un tipo de riesgo para el que se ofertan coberturas. Los supuestos fundamentales de este modelo son la independencia entre las pólizas, y por consiguiente entre la exposición al riesgo de cada persona u objeto cubierto por las mismas, y que durante el periodo de análisis las pólizas pueden no presentar reclamaciones o una única reclamación.

La primera suposición puede ser argumentada dado que existen circunstancias en las que la ocurrencia de siniestros entre pólizas distintas pueden estar correlacionada, por ejemplo si se toma una compañía de seguros AA que emite pólizas de cobertura al riesgo automovilístico para la mitad de la población, es muy probable que si sucede una carambola más de una póliza cubierta se vea implicada en el mismo accidente. Sin embargo, los argumentos en contra de esta suposición son intrascendentes y por lo tanto ignorados.

La segunda suposición es un tanto mas delicada ya que con excepción del seguro de vida, los riesgos cubiertos pueden ocurrir más de una vez en el periodo de observación, que en el ámbito asegurador generalmente es un año. Una forma de forzar a que la suposición sea verdadera es seleccionando un intervalo de tiempo lo suficientemente pequeño para que a lo más pueda ocurrir un siniestro por póliza.

Relacionado al problema que enfrenta esta suposición, es que surge el segundo modelo que se mencionaba y que es comúnmente conocido como el *Modelo Agregado*, en el cual se considera una póliza o portafolio de pólizas donde el número de reclamaciones es aleatorio con lo que existen dos fuentes de variabilidad en el modelo, la del número de reclamaciones y la del monto de cada reclamación individual.

Como se puede ver, los dos modelos son muy distintos desde su concepción y naturaleza por lo que el desarrollo matemático de cada uno también es muy diferente. A continuación se presentarán ambos desarrollos y sus resultados.

### 3.1. Modelo Individual

Como se dijo anteriormente el modelo individual tiene su base en una cartera con un número de pólizas o riesgos fijos representado por  $n$ , donde cada uno de estos riesgos es desconocido e independiente de los otros. Sean los montos de reclamación de la cartera de pólizas una serie de variables aleatorias  $X_1, X_2, \dots, X_n$  tal que  $x_j$  es el monto de reclamo de la v.a.  $X_j$ .

Es importante que aunque las variables se suponen independientes entre si, en ningún momento se supone que tengan alguna distribución común por lo que este modelo es aplicable a carteras que contengan riesgos no necesariamente homogéneos.

Del segundo supuesto en el que se basa el modelo individual se entiende que existe una probabilidad, digamos  $q_j$  de que la póliza  $j$ -ésima presente alguna reclamación; por lo que aunado a la variable  $X_j$ , para cada  $j$  entre 1 y  $n$ , existirá una variable aleatoria Bernoulli de reclamación  $R_j$  tal que,

$$R_j = \begin{cases} 1 & \text{con probabilidad } q_j \\ 0 & \text{con probabilidad } 1 - q_j \end{cases}$$

Para describir el monto de reclamación de una póliza es necesario considerar ambas variables ya que por ejemplo, la probabilidad de que el monto sea igual a algún valor de interés dependerá en gran medida de la probabilidad de que exista dicho reclamo. Por lo tanto para poder describir las propiedades del modelo es necesario expresar la relación entre las variables; a continuación se listan los supuestos de independencia entre ellas:

- Las v.a.  $R_j$  son independientes entre si, i.e.  $R_i \perp R_j \quad \forall i \neq j$
- Las v.a.  $X_j$  son independientes entre si, i.e.  $X_i \perp X_j \quad \forall i \neq j$
- Las v.a.  $R_j$  son independientes de las variables  $X_j$ , i.e.  $X_j \perp R_j$

Así, defínase  $S_j$  como la variable del riesgo  $j$ -ésimo tal que  $S_j = R_j X_j$ , entonces la variable estará dada como,

$$S_j = R_j X_j \begin{cases} X_j & \text{si } R_j = 1 \\ 0 & \text{si } R_j = 0 \end{cases}$$

Por lo que el monto total o *agregado* de reclamación será la suma de los montos individuales de las  $n$  pólizas o riesgos.

$$S = \sum_{j=1}^n S_j \quad (3.1)$$

Denotamos por  $F(x)$  la función de distribución del monto total, por  $F_j(x)$  la función de distribución del riesgo individual  $S_j$  y por  $G(x)$  la función de distribución del monto de reclamación. Entonces,<sup>1</sup>

$$F(x) = (F_1 * F_2 * \dots * F_n)(x)$$

De lo cual siguen las siguientes proposiciones:

1.  $F_j(x) = 1 - q_j(1 - G_j(x)) \quad \text{si } x \leq 0$
2.  $\mathbb{E}[S] = \sum_{j=1}^n q_j \mathbb{E}[x_j]$
3.  $Var[S] = \sum_{j=1}^n q_j Var[X_j] + (1 - q_j) \mathbb{E}[X_j]^2$
4.  $m_{S_j}(t) = 1 - q_j(1 - m_{X_j}(t))$
5.  $m_S(t) = \prod_{j=1}^n 1 - q_j(1 - m_{X_j}(t))$

A continuación se presentan las demostraciones respectivas.

---

<sup>1</sup>Se representa la operación de convolución mediante el símbolo \*.

**Demostración. 1**

$$\begin{aligned}
F_j(x) &= \mathbb{P}[S_j \leq x] \\
&= \mathbb{P}[R_j X_j \leq x] \\
&= \mathbb{P}[R_j X_j \leq x | R_j = 0] \mathbb{P}[R_j = 0] + \mathbb{P}[R_j X_j \leq x | R_j = 1] \mathbb{P}[R_j = 1] \\
&= 1 - q_j + q_j g_j(x) \\
\text{Por lo tanto, } F_j(x) &= 1 - q_j(1 - g_j(x)) \quad \blacksquare
\end{aligned}$$

**Demostración. 2**

$$\begin{aligned}
\mathbb{E}[S] &= \mathbb{E}\left[\sum_{j=1}^n S_j\right] \\
&= \sum_{j=1}^n \mathbb{E}[R_j X_j] \\
&= \sum_{j=1}^n \mathbb{E}[R_j] \mathbb{E}[X_j]
\end{aligned}$$

$$\text{Por lo tanto, } \mathbb{E}[S] = \sum_{j=1}^n q_j \mathbb{E}[x_j] \quad \blacksquare$$

**Demostración. 3**

$$\begin{aligned}
\text{Var}[S_j] &= \mathbb{E}[R_j X_j^2] - \mathbb{E}[R_j X_j]^2 \\
&= \mathbb{E}[R_j^2] \mathbb{E}[X_j^2] - \mathbb{E}[x_j]^2 \mathbb{E}[X_j]^2 \\
&= q_j \mathbb{E}[X_j^2] - q_j^2 \mathbb{E}[X_j]^2 \\
&= q_j (\mathbb{E}[x_j^2] - \mathbb{E}[X_j]^2 + \mathbb{E}[X_j]^2 - q_j \mathbb{E}[X_j]^2) \\
&= q_j (\text{Var}[x_j] + (1 - q_j) \mathbb{E}[X_j]^2)
\end{aligned}$$

Por otra parte:

$$\begin{aligned}
\text{Var}[S] &= \text{Var}\left[\sum_{i=1}^n S_j\right] \\
&= \sum_{j=1}^n \text{Var}[S_j]
\end{aligned}$$

$$\text{Por lo tanto, } \text{Var}[S] = \sum_{i=1}^n q_j \text{Var}[X_j] + (1 - q_j) \mathbb{E}[X_j]^2 \quad \blacksquare$$

**Demostración. 4**

$$\begin{aligned}
m_{s_j}(t) &= \mathbb{E}[e^{tR_j X_j}] \\
&= \mathbb{E}[e^{tR_j X_j} | R_j = 0] \mathbb{P}[R_j = 0] + \mathbb{E}[e^{tR_j X_j} | R_j = 1] \mathbb{P}[R_j = 1] \\
&= 1 - q_j + q_j m_{X_j}(t) \\
\text{Por lo tanto, } m_{S_j}(t) &= 1 - q_j(1 - m_{X_j}(t)) \quad \blacksquare
\end{aligned}$$



**Demostración. 5**

$$\begin{aligned}
m_S(t) &= \mathbb{E}[e^{tS}] \\
&= \mathbb{E}[e^{t\sum_{j=1}^n S_j}] \\
&= \mathbb{E}\left[\prod_{j=1}^n e^{tS_j}\right] \\
&= \prod_{j=1}^n \mathbb{E}[e^{tS_j}]
\end{aligned}$$

$$\text{Por lo tanto, } m_S(t) = \prod_{i=1}^n 1 - q_j(1 - m_{X_j}(t)) \quad \blacksquare$$

### 3.2. Modelo Compuesto

El modelo agregado o compuesto hace frente al supuesto de número de reclamaciones del modelo individual, que lo limita a presentar únicamente entre cero y  $n$  reclamaciones, donde  $n$  es un valor fijo. Este modelo permite mayor flexibilidad en este sentido ya que se consideran reclamaciones variables, ya sea para una única póliza o para un conjunto de pólizas.<sup>2</sup>

Esto además hace al modelo un tanto más realista ya que una empresa aseguradora, al inicio del periodo de observación de sus coberturas no sabe cuántas reclamaciones puedan ocurrir ni, en caso de que se presenten, de qué monto será cada una de ellas.

Sea  $N$  la variable aleatoria que representa el número de reclamaciones existentes durante el periodo de observación y  $\{X_j\}_{j \in \mathbb{N}}$  la sucesión de variables aleatorias no negativas que representa los montos de reclamación. Al igual que en el modelo individual, donde el monto de reclamación total estaba dado por la información aportada por la variable de monto y por la de ocurrencia, en este modelo la relación entre las variables  $\{X_j\}_{j \in \mathbb{N}}$  y  $N$  es la base para desarrollar el modelo. Así pues, a continuación se presentan los supuestos sobre los que se construye el modelo:

1.  $\{X_j\}_{j \in \mathbb{N}}$  es una sucesión de variables aleatorias independientes idénticamente distribuidas.
2. La distribución de las variables  $X_1, X_2, \dots, X_j, \dots, X_n$  no depende de  $N$ .
3. La distribución de  $N$  no depende de  $X_1, X_2, \dots, X_j, \dots, X_n$ .

El primer supuesto indica homogeneidad en los riesgos tratados. Esta es la segunda gran diferencia conceptual respecto al modelo individual, ya que en este último

---

<sup>2</sup>A pesar de que el modelo es aplicable tanto para modelar el costo total de una póliza que permita múltiples reclamaciones o un conjunto de pólizas similares, la interpretación usual es como la severidad total de un portafolio de riesgos homogéneos.

no se especificaba alguna condición sobre la distribución de las variables de monto, pudiendo ser que cada variable tuviera una distribución diferente.

Además, la independencia a la que refiere el primer supuesto, significa que el monto de una reclamación no depende del monto de ninguna otra reclamación, mientras que la independencia referida en los supuestos dos y tres, dice que el número de reclamaciones no tiene ningún efecto sobre el monto de las reclamaciones y viceversa.

Así pues, el monto total de reclamaciones de los riesgos durante el periodo de observación está dado por la suma de los montos de reclamación individuales, que expresados en términos de la variable agregada, se plantea como sigue:

$$S = \sum_{i=1}^N X_i. \quad (3.2)$$

De lo anterior se sigue que para que el monto total de reclamación sea igual a cero se necesita que no se haya presentado ninguna reclamación, i.e.  $N = 0$  por lo cual se excluye el cero del soporte de las variables  $X_j$  y para calcular la probabilidad que  $S = 0$  basta con calcular  $P[N = 0]$ .

Algunas de las características del modelo agregado son las siguientes:

1.  $\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[X]$
2.  $\mathbb{E}[S^2] = \mathbb{E}[N]\mathbb{E}[X^2] + \mathbb{E}[N(N-1)]\mathbb{E}[X]^2$
3.  $Var[S] = \mathbb{E}[N]Var[X] + \mathbb{E}[X^2]Var[N]$
4.  $m_S(t) = m_N(\log(m_x(t)))$
5.  $f_S(s) = \sum_{n=0}^{\infty} f^{*n}(x)P[N = n]$

### **Demostración. 1**

$$\begin{aligned} \mathbb{E}[S] &= \mathbb{E}\left[\sum_{i=1}^N X_i\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^N X_j | N = n\right] P[N = n] \\ &= \sum_{n=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^n X_j\right] P[N = n] \\ &= \sum_{n=0}^{\infty} \sum_{i=1}^n \mathbb{E}[X_j] P[N = n] \\ &= \sum_{n=0}^{\infty} n \mathbb{E}[X] P[N = n] \\ &= \mathbb{E}[X] \sum_{n=0}^{\infty} n P[N = n] \end{aligned}$$

Por lo tanto,  $\mathbb{E}[S] = \mathbb{E}[X]\mathbb{E}[N]$  ■

Lo cual resulta lógico, ya que se está reflejando que el monto total esperado está dado por el monto esperado de cada reclamación individual, multiplicado por la cantidad de veces que se espera que se materialice el riesgo y por tanto, exista una reclamación.

**Demostración. 2**

$$\begin{aligned}
\mathbb{E}[S^2] &= \mathbb{E}\left[\left(\sum_{i=1}^N X_i\right)^2\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^2\right] \mathbb{P}[N = n] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^n X_i^2 + 2 \sum_{i < j} X_i X_j\right] \mathbb{P}[N = n] \\
&= \sum_{n=0}^{\infty} \left(n \mathbb{E}[X^2] + 2 \left(\frac{(n-1)n}{2}\right) \mathbb{E}[X]^2\right) \mathbb{P}[N = n] \\
&= \mathbb{E}[X^2] \sum_{n=0}^{\infty} n \mathbb{P}[N = n] + \mathbb{E}[X]^2 \sum_{n=0}^{\infty} (n-1)n \mathbb{P}[N = n] \\
\text{Por lo tanto, } \mathbb{E}[S^2] &= \mathbb{E}[X^2] \mathbb{E}[N] + \mathbb{E}[X]^2 \mathbb{E}[N(N-1)] \quad \blacksquare
\end{aligned}$$

**Demostración. 3**

$$\begin{aligned}
\text{Var}[S] &= \mathbb{E}[S^2] - \mathbb{E}[S]^2 \\
&= \mathbb{E}[X^2] \mathbb{E}[N] + \mathbb{E}[X]^2 \mathbb{E}[N(N-1)] - \mathbb{E}[X]^2 \mathbb{E}[N]^2 \\
&= \mathbb{E}[X^2] \mathbb{E}[N] + \mathbb{E}[X]^2 (\mathbb{E}[N^2] - \mathbb{E}[N]) - \mathbb{E}[X]^2 \mathbb{E}[N]^2 \\
&= \mathbb{E}[X^2] \mathbb{E}[N] - \mathbb{E}[X]^2 \mathbb{E}[N] + \mathbb{E}[X]^2 \mathbb{E}[N^2] - \mathbb{E}[X]^2 \mathbb{E}[N]^2 \\
\text{Por lo tanto, } \text{Var}[S] &= \mathbb{E}[N] \text{Var}[X] + \mathbb{E}[X]^2 \text{Var}[N] \quad \blacksquare
\end{aligned}$$

La varianza de  $S$  también puede obtenerse por medio de la varianza iterada, así:

$$\begin{aligned}
\text{Var}[S] &= \mathbb{E}[\text{Var}[S|N]] + \text{Var}[\mathbb{E}[S|N]] \\
&= \mathbb{E}\left[\text{Var}\left[\sum_{i=1}^N X_i | N\right]\right] + \text{Var}\left[\mathbb{E}\left[\sum_{i=1}^N X_i | N\right]\right] \\
&= \mathbb{E}[N \text{Var}[X]] + \text{Var}[N \mathbb{E}[X]] \\
\text{Por lo tanto, } \text{Var}[S] &= \mathbb{E}[N] \text{Var}[X] + \mathbb{E}[X]^2 \text{Var}[N] \quad \blacksquare
\end{aligned}$$

**Demostración. 4**

$$\begin{aligned}
m_S(t) &= \mathbb{E}[e^{tS}] = \mathbb{E}[e^{t\sum_{i=1}^N X_i}] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[e^{t\sum_{i=1}^N X_i} | N = n] \mathbb{P}[N = n] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[e^{t\sum_{i=1}^n X_i}] \mathbb{P}[N = n] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[\prod_{i=1}^n e^{tX_i}\right] \mathbb{P}[N = n] \\
&= \sum_{n=0}^{\infty} \prod_{i=1}^n \mathbb{E}[e^{tX_i}] \mathbb{P}[N = n] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[e^{tX}]^n \mathbb{P}[N = n] \\
&= \sum_{n=0}^{\infty} m_X(t)^n \mathbb{P}[N = n] \\
&= \mathbb{E}[m_X(t)^N] = \mathbb{E}[e^{N \ln(m_X(t))}] \\
&\text{Por lo tanto, } m_S(t) = m_N(\ln(m_X(t))) \quad \blacksquare
\end{aligned}$$

**Demostración. 5**

Sabemos que la distribución de la pérdida agregada está dada por:<sup>3</sup>

$$F_S(s) = P(S \leq s) = \sum_{n=0}^{\infty} \mathbb{P}[S \leq s | N = n] \mathbb{P}[N = n].$$

Sea  $F_X(x) = P(X \leq x)$  la función de distribución común de las variables  $X_j$  y denotemos  $P(N = n)$  como  $p_n$ , entonces

$$F_S(s) = \sum_{n=0}^{\infty} F_X^{*n}(x) p_n,$$

donde  $F_X^{*n}(x)$  representa la  $n$ -ésima convolución de la distribución de  $X$ , la cual puede ser obtenida de forma recursiva como:

$$F_X^{*0}(x) \begin{cases} 0, & x < 0 \\ 1, & x \leq 0 \end{cases}$$

$$F_X^{*k}(x) = \int_{-\infty}^{\infty} F_X^{*k-1}(x)(x-y) dF_X(y) \quad \text{para } k = 1, 2, \dots$$

<sup>3</sup>Stuart A. Klugman, Harry H. Panjer, Gordon E. Willmot. (2004). Aggregate loss models. En Loss Models: From Data to Decisions(140-142). New Jersey, Estados Unidos: John Wiley & Sons

Si  $X$  es una variable aleatoria continua con soporte en valores no negativos, la convolución se reduce a:

$$F_X^{*k}(x) = \int_0^x F_X^{*(k-1)}(x-y)f_X(y)dy \quad \text{para } k = 1, 2, \dots$$

Mientras que si la variable es discreta con soporte en  $0, 1, 2, \dots$ , se reduce a:

$$F_X^{*k}(x) = \sum_{y=0}^x F_X^{*(k-1)}(x-y)f_X(y) \quad \text{para } k = 1, 2, \dots$$

Por lo que la función de probabilidad correspondiente sería:

$$f_X^{*k}(x) = \sum_{y=0}^x f_X^{*(k-1)}(x-y)f_X(y) \quad \text{para } k = 1, 2, \dots$$

Por lo tanto, la función de densidad de las pérdidas agregadas es:

$$f_S(s) = \sum_{n=0}^{\infty} f^{*n}(x)p_n \quad \blacksquare \quad (3.3)$$

El objetivo final de modelar un riesgo es encontrar una forma de describir lo mejor posible las características que lo conforman, en el sentido del seguro esto significa hallar la función de distribución que mejor ajuste a las pérdidas agregadas, sin embargo, hallar una distribución exacta para  $S$  es complicado particularmente en situaciones donde la cantidad de información es muy grande. Por lo tanto, generalmente se modela por separado la distribución de  $N$  y la distribución común  $X$ ; este enfoque tiene varias ventajas tales como:

- Si el análisis se realiza sobre una cartera de riesgos de una compañía en crecimiento, el hecho de tener modelada la frecuencia por separado hace que la actualización de los datos y el consecuente análisis sea más práctico.
- Permite reflejar de mejor manera sobre el modelo agregado los efectos de diversos factores externos, ya que identificar los efectos de dichos factores sobre un modelo general es más complicado que si se analiza la severidad y frecuencia por separado. Por ejemplo, en el caso económico o financiero, sería más fácil incorporar el efecto de la inflación e imposición de tarifas o normas fiscales.
- En el entorno asegurador este enfoque permite observar el impacto de medidas de corresponsabilidad de riesgos como los deducibles, coaseguros y límites de póliza y por tanto, generar mejores indicadores para determinar la viabilidad de productos nuevos o la rentabilidad de productos existentes.
- Los modelos desarrollados para calcular pérdidas para las empresas de seguro y para calcular reclamaciones a reaseguradoras suelen ser más compatibles permitiendo a ambos sectores tener mayor facilidad de planeación financiera.

- La distribución de  $S$  estará determinada por las distribuciones de la frecuencia  $N$  y la severidad  $X$  por lo que entender la información relacionada a cada una permite un mejor entendimiento de la información general.

Por todas las consideraciones que se han presentado sobre modelo agregado puede vislumbrarse la complejidad y las distintas formas que puede tomar el modelo. Por ello es que se presentan algunas formas del modelo con mayor profundidad y siguiendo la nomenclatura usual, es decir, brindando al modelo de características particulares fijando la distribución de la variable de frecuencia.

### Modelo Poisson Compuesto

El modelo Poisson Compuesto es seguramente el modelo más utilizado por su extendido estudio en el campo de los procesos estocásticos en donde tiene su origen. En sí, el proceso Poisson es un proceso a tiempo continuo donde tanto los saltos como su tamaño son variables tal que:

$$Y(t) = \sum_{i=1}^{N(t)} X_i \quad N(t) \sim \text{Poisson}(\lambda t).$$

Los usos de este proceso se extienden a la Biología mediante el estudio del comportamiento de ecosistemas, la Física en el análisis de la distribución de radiación procedente de sistemas no estables, investigación de operaciones para la optimización de procesos de filas, finanzas para estimación y prevención de ruina, etc.<sup>4</sup>

Los campos donde pueden encontrarse aplicaciones para el modelo aumentaron exponencialmente mediante la inclusión de variabilidad en el parámetro de la distribución Poisson relacionada a la frecuencia, lo que dio paso al *Proceso Poisson No Homogéneo*; ya que permite estudiar fenómenos donde el proceso de llegada no es uniforme en el tiempo, sino que sigue cierto comportamiento respecto al tiempo.

En el caso del aseguramiento y el análisis de riesgos financieros en instituciones bancarias, la base utilizada es el proceso Poisson homogéneo donde el tiempo está dado por una unidad, generalmente anual y por tanto es sobre el que se profundizará a continuación mostrando algunas de sus propiedades más importantes.

$$\begin{aligned} \mathbb{E}[S] &= \lambda \mathbb{E}[X], \\ \text{Var}[S] &= \lambda \mathbb{E}[X^2], \\ m_S(t) &= e^{\lambda(m_X(t)-1)}. \end{aligned}$$

---

<sup>4</sup>Properties of compound poisson processes with applications in statistical physics, consulta 02/02/2018, fuente: <https://www.sciencedirect.com/science/article/pii/0031891469900962>

### Modelo Binomial Compuesto

Del mismo modo que el modelo Poisson Compuesto adquiere su nombre de la distribución asignada a su frecuencia, así también sucede en el modelo Binomial Compuesto. Por tanto, si la frecuencia  $N$  tiene distribución Binomial y la severidad  $X$  tiene una distribución  $F_X$  con soporte en los reales no negativos, se dice que  $S$  tiene una distribución “Binomial Compuesta”.

Las características de este modelo también derivan de las cuatro primeras propiedades del modelo agregado general y son,

$$\begin{aligned}\mathbb{E}[S] &= np\mathbb{E}[X], \\ \text{Var}[S] &= np\mathbb{E}[X^2] - np^2\mathbb{E}[X]^2, \\ m_S(t) &= (1 - p + pm_X(t))^n.\end{aligned}$$

### Modelo Binomial Negativo Compuesto

Si el número de reclamaciones  $N$  se distribuye Binomial Negativa, y se cumplen las características requeridas sobre la variable  $X$  de severidad, entonces se dice que el modelo agregado  $S$  tiene una distribución Binomial Negativa Compuesta y tiene las siguientes características,

$$\begin{aligned}\mathbb{E}[S] &= r(p^{-1} - 1)\mathbb{E}[X], \\ \text{Var}[S] &= r(p^{-1} - 1)\mathbb{E}[X^2] - r(p^{-1} - 1)^2\mathbb{E}[X]^2, \\ m_S(t) &= \left( \frac{p}{1 - (1 - p)m_X(t)} \right)^r.\end{aligned}$$

## 3.3. Aproximaciones

Hemos presentado una pequeña parte de la teoría detrás del modelo agregado  $S$ , así como algunas de las problemáticas que se pueden llegar a enfrentar al momento de realizarlo. Una manera de evitar las dificultades inherentes del modelo es el ajuste de aproximaciones. Por ello, en esta sección se presentan de manera breve tres aproximaciones: Normal, Log-normal y Gamma.

### 3.3.1. Aproximación Normal

La distribución Normal o *Gaussiana* es una de las distribuciones más conocidas de la Estadística y una de las utilizadas con mayor frecuencia gracias a su forma y propiedades características que la hacen altamente práctica en diversas áreas del conocimiento.

Recordamos pues que para  $X \sim N(\mu, \sigma^2)$ :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad \mu \in \mathbb{R}, \sigma^2 > 0.$$

Y en particular la v.a.  $Z$  sigue una distribución *Normal estándar* si:

$$Z = \frac{X - \mu}{\sigma} \sim N(\mu = 0, \sigma^2 = 1).$$

La función  $\phi$  es ampliamente conocida y representa la probabilidad de que  $Z$  sea menor a un valor  $x$ , es decir

$$\phi(x) = P[Z \leq x] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Considerando el Teorema central del límite, podemos ver la aproximación al modelo dada una  $s$  en el soporte de  $S$  como:

$$F_S(s) = P[S \leq s] \approx \phi\left(\frac{s - \mathbb{E}[S]}{\sqrt{Var[S]}}\right). \quad (3.4)$$

Uno de los dos principales problemas que presenta esta aproximación es que, como la distribución Normal es simétrica respecto a su media, es un mal ajuste a modelos agregados que presentan colas pesadas, porque sobrestimaría eventos con valores circundantes a  $\mathbb{E}[S]$  en perjuicio de valores más extremos.

El segundo problema es que al contrario del modelo agregado que únicamente toma valores en los reales no negativos, esta aproximación sí puede tomar valores negativos.

Una de las formas más comunes de vadear ese problema es tener en consideración la regla empírica de la distribución Normal que, aunque sus autores Murray R. Spiegel y Larry J. Stephens no la propusieron con esa finalidad, fueron los primeros en asentarla por escrito en su trabajo publicado en 1989.<sup>5</sup>

La regla empírica o regla “68-95-99”, establece que para una v.a  $X$  con distribución Normal, aproximadamente el 68 % de la probabilidad se acumula al rededor de la media a distancia no mayor de una desviación estándar, de manera similar el 95 % y 99 % de la probabilidad se acumula al rededor de la media mas 2 y 3 desviaciones estándar.<sup>6</sup>

De manera matemática la regla se expresa como:

$$\begin{aligned} P[X \in (\mu \pm \sigma)] &\approx 0.6827, \\ P[X \in (\mu \pm 2\sigma)] &\approx 0.9545, \\ P[X \in (\mu \pm 3\sigma)] &\approx 0.9973. \end{aligned}$$

<sup>5</sup>Murray R. Spiegel, Larry J. Stephens. (1989). The Standard Daviation and Other Measures of Dispersion. En Schaum’s Outline of Theory and Problems of Statistics (95-101). Estados Unidos: McGraw-Hill.

<sup>6</sup>Las cifras que se manejan en el libro son 68.27, 95.45 y 99.73 respectivamente



Y para una variable con distribución Normal estándar sería:

$$P[Z \in (-1, 1)] \approx 0.6827,$$

$$P[Z \in (-2, 2)] \approx 0.9545,$$

$$P[Z \in (-3, 3)] \approx 0.9973.$$

Así, la práctica común es tomar los valores de media y desviación de la muestra de datos y observar ya sea analítica o gráficamente si el valor cero está a dos, tres o más desviaciones estándar de la media; si es así, se considera que los datos negativos que se pudieran generar de la simulación de la aproximación no aportan información significativa y pueden ser ignorados. Naturalmente esta práctica, como muchas otras dentro de la estadística clásica es altamente subjetiva pero útil en ciertas ocasiones.

### 3.3.2. Aproximación Log-normal

Como medida para solucionar los problemas de la aproximación anterior, se puede optar por buscar una aproximación a  $S$  mediante la distribución Log-normal pues tiene una forma y soporte más apropiados.

Así pues, si una variable  $Y$  se distribuye Normal con parámetros  $\mu$  y  $\sigma$ , entonces la v.a.  $X$  se distribuye *Log-normal* si  $X = \exp(Y)$  y se tiene que:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(\ln(x)-\mu)^2/2\sigma^2} \quad x > 0 \quad \mu \in \mathbb{R}, \sigma^2 > 0$$

El valor de la esperanza es  $\mathbb{E}[X] = e^{\mu+\sigma^2/2}$ , mientras que la varianza está dada por  $Var(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$ .

De manera similar a la aproximación Normal del modelo agregado, podemos ver esta nueva aproximación dada una  $s > 0$  como:

$$F_S(s) = P[S \leq s] \approx \phi\left(\frac{\ln(s) - \mathbb{E}[S]}{\sqrt{Var[S]}}\right) \quad (3.5)$$

A pesar de que esta distribución cumple particularidades similares a las que cumplen los datos de un modelo de riesgos agregados, no existe forma de decir de antemano qué tan bueno será el ajuste de esta distribución. Lo anterior es principalmente por que, al ser una aproximación, la “bondad de ajuste” al modelo es más subjetiva y depende del uso que quien lo analiza quiera darle.

### 3.3.3. Aproximación Gamma trasladada

La construcción de esta aproximación difiere de las dos anteriores, puesto que se sigue una guía similar a la estimación paramétrica por momentos. Así pues, supongamos un valor constante  $x_0$  y una variable aleatoria  $X$  con distribución  $Gamma(\alpha, \theta)$  y función de densidad,

$$f_X(x|\alpha, \theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} \quad \text{con } \alpha \geq 0 \quad \text{y } \theta > 0.$$

La función de distribución de la variable Gamma trasladada  $x_0 + X$  está dada por:

$$G^{tr}(x|\alpha, \theta, x_0) = F(x - x_0|\alpha, \theta).$$

Para realizar la aproximación, los parámetros  $\alpha$ ,  $\theta$  y  $x_0$  tendrán que ser tales que el primer, segundo y tercer momento central sean iguales a los parámetros correspondientes a la distribución Gamma trasladada a estimar.<sup>7</sup>

De tal modo, se supondrán conocidos los valores de los primeros tres momentos de  $S$ , siendo éstos  $\mathbb{E}[S] = \mu$ ,  $Var[S] = \sigma^2$  y  $\frac{\mathbb{E}[(S - \mathbb{E}[S])^3]}{Var(S)^{3/2}} = \tau$  con  $\tau > 0$  para denotar sesgo positivo tal como en la ecuación (1.2).

Para la variable  $x_0 + X$  se tiene:

- $\mathbb{E}[x_0 + X] = x_0 + \alpha\theta.$
- $Var[x_0 + X] = \alpha\theta^2.$
- $\frac{\mathbb{E}[(x_0 + X - \mathbb{E}[x_0 + X])^3]}{Var(x_0 + X)^{3/2}} = \frac{2}{\sqrt{\alpha}}.$

De lo anterior se obtiene el siguiente sistema de ecuaciones:

$$\begin{aligned} x_0 + \alpha\theta &= \mu, \\ \alpha\theta^2 &= \sigma^2, \\ \frac{2}{\sqrt{\alpha}} &= \tau. \end{aligned} \tag{3.6}$$

Por lo tanto, la distribución aproximada de  $S$  estará dada por los parámetros  $\hat{x}_0 = \mu - \frac{2}{\tau}$ ,  $\hat{\alpha} = \frac{4}{\tau^2}$  y  $\hat{\theta} = \frac{\sigma\tau}{2}$  con función de distribución aproximada:

$$G_S^{tr}(s|\alpha, \theta, x_0) \approx F\left(s - \left(\mu - \frac{2}{\tau}\right) \middle| \alpha = \frac{4}{\tau^2}, \theta = \frac{\sigma\tau}{2}\right). \tag{3.7}$$

---

<sup>7</sup>Pavel Cížek, Wolfgang Härdle, Rafal Weron . (2005). Statistical Tools for Finance and Insurance. Alemania: Humboldt Universität.

# Capítulo 4

## Ajuste y selección del modelo

### 4.1. Presentación de los datos

Los datos utilizados en el análisis para este trabajo fueron obtenidos de la página oficial del Aeropuerto Internacional de la Ciudad de México (AICM) en forma de registros diarios del 12 de junio del 2016 al 5 de noviembre del mismo año. Los registros están conformados por 5 columnas, estas son: el día de llegada al aeropuerto, la clave de la compañía aérea, el código o número del vuelo, la hora registrada de llegada y la diferencia en minutos enteros entre la hora programada de arribo y la hora real de llegada.

A continuación se hace el listado de las aerolíneas que tuvieron operación en el AICM durante el período de estudio, así como su respectiva clave. En la tabla 4.1 se presentan las 7 aerolíneas nacionales y en la tabla 4.2 las 23 aerolíneas de procedencia extranjera.

Nombre	Clave
Aeromar	TAO
Aeromexico	AMX
Aeromexico Connect	SLI
Interjet	AIJ
Magnicharters	GMT
Vivaaerobus	VIV
Volaris	VOI

Tabla 4.1: Aerolíneas nacionales

Nombre	Clave
Aero República	RPB
Air Canada	ACA
Air France	AFR
Alaska Airlines	ASA
Southwest Airlines Co	SWA
American Airlines	AAL
Avianca	AVA
British Airlines	BAW
Copa Airlines	CPM
Cubana de Aviación	CUB
Delta Airlines	DAL
Iberia	IBE
JetBlue	JBU
KLM	KLM
Lacsa	LRC
Latam Airlines Group S.A.	LAM
TAM Linhas Aereas S.A.	TAM
Lufthansa	DLH
Taca	TAI
United Airlines	UAL
Alitalia	AZA
All Nippon Airways CO.	ANA
China Southern Airlines	CSN

Tabla 4.2: Aerolíneas extranjeras

La base de datos recopilada consta de un total de 73,389 vuelos con incumplimiento registrado, de los cuales 65,089 fueron operados por aerolíneas nacionales y 8,300 por aerolíneas internacionales<sup>1</sup>, es decir, únicamente 11.3% de los vuelos que llegaron con retraso al AICM durante el periodo de análisis fueron pilotados por aerolíneas listadas en la tabla 4.2.

Es importante destacar que en los registros no se incluye la totalidad de llegadas al aeropuerto, sino que se registran aquellos vuelos cuyo retraso superó los 15 minutos.

<sup>1</sup>La clasificación de las aerolíneas se realiza separando las aerolíneas que fueron fundadas en territorio nacional, y por tanto son mexicanas dentro del ámbito legal, de las aerolíneas con orígenes extranjeros. Nótese que no excluye en ningún momento la posibilidad que aerolíneas mexicanas puedan operar vuelos con origen o destino internacional por lo que, el decir *aerolíneas internacionales*, se hace con términos nominativos refiriéndose a la segunda categoría de la clasificación.

El número total de arribos al aeropuerto de vuelos operados por aerolíneas nacionales durante el período de observación fue de 126,316 por lo que el número de vuelos que llegaron dentro del margen considerado como llegada puntual fue de 61,227, es decir, la proporción de vuelos con retrasos registrados como incumplimiento en tiempo es de 0.515287.

En el caso de los que fueron operados por las aerolíneas extranjeras, el total de llegadas al AICM fue de 46,074 vuelos, de los cuales 37,774 llegaron dentro de límite considerado como llegada a tiempo, por lo que la probabilidad de registrar incumplimiento es de 0.180145.

Teniendo en cuenta lo anterior, se presenta el análisis descriptivo de la frecuencia. Para obtener el vector de frecuencia se tienen que contabilizar los eventos ocurridos en los intervalos en que se ha dividido el plazo de tiempo observado, en este caso serán los 146 días en la base de datos.

De lo anterior se obtienen dos vectores, uno de frecuencia para vuelos operados por aerolíneas nacionales y uno segundo para aquellos bajo aerolíneas internacionales.

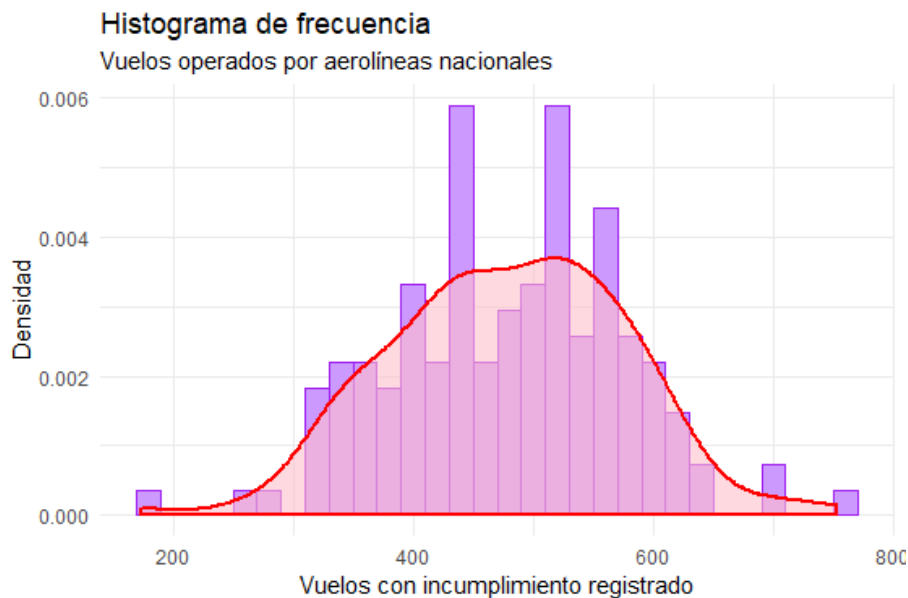


Figura 4.1: Histograma de frecuencia de aerolíneas nacionales

Como se puede apreciar, la densidad de ambos vectores toma valores pequeños, ya que los datos tienen soportes en intervalos amplios. Otro factor interesante presente en las gráficas es que en ambos casos, en la campana de la distribución, existe un rango de valores con frecuencia menor a los valores circundantes; es más notorio en el histograma de la frecuencia para aerolíneas extranjeras en el cual, la sexta barra que cubre valores cercanos a 60 es menor que las dos barras aledañas.

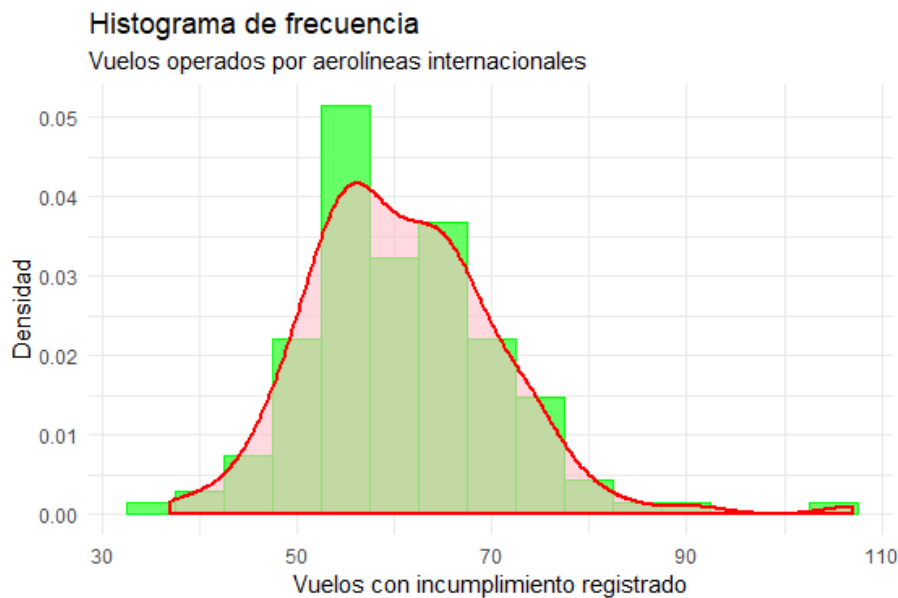


Figura 4.2: Histograma de frecuencia de aerolíneas internacionales

Una de las diferencias principales entre los vectores es el sesgo que presentan. En la gráfica 4.1, se muestra un ligero sesgo negativo mientras que en la segunda el sesgo es positivo.

Para complementar el análisis gráfico de los vectores de frecuencia, se muestra en la tabla 4.3 algunas métricas estadísticas básicas de los datos en conjunto y para los datos ya clasificados en vuelos operados por aerolíneas nacionales e internacionales.

	Agregados	Nacionales	Internacionales
Media	539.6	478.6	61.02206
Mediana	541.0	485.5	60.00
Varianza	10618.89	9354.777	98.15507
SD	103.048	96.7201	9.907324

Tabla 4.3: Valores de media, varianza y desviación estándar de frecuencia

A continuación, en la imagen 4.3 se presentan gráficamente los vectores que clasifican la severidad de la misma forma que los vectores anteriores clasifican la frecuencia.

La forma de la función de densidad de ambos vectores de severidad es en extremo similar, sin embargo, se puede apreciar que el tiempo de retraso para vuelos operados por aerolíneas nacionales presenta colas más pesadas que aquellos operados internacionalmente.

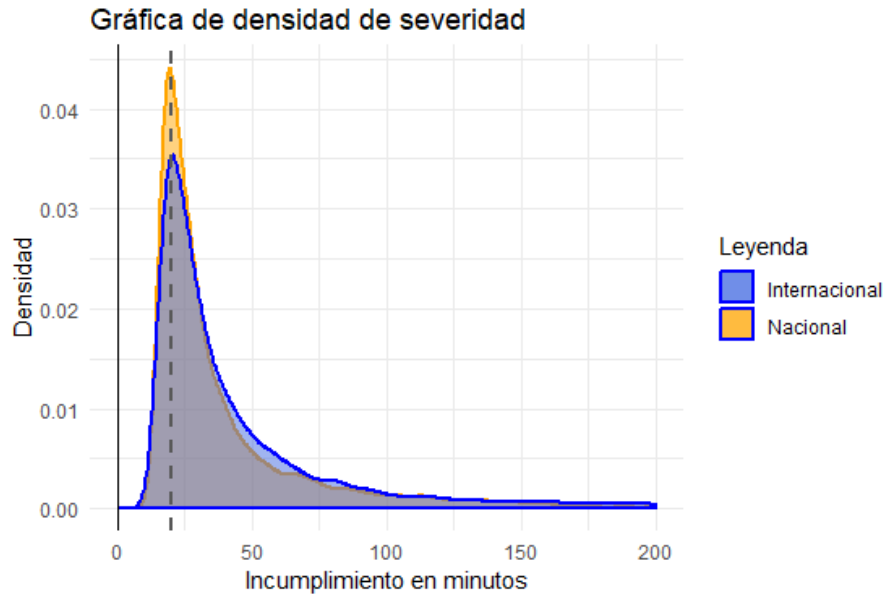


Figura 4.3: Densidad de los vectores de severidad

Los intervalos que abarcan los valores de severidad de las muestras de los riesgos analizados son bastante amplios, por lo que una herramienta útil para el análisis de características más detalladas de sus densidades es el reescalamiento. Una función comúnmente utilizada para ese propósito es el logaritmo ya que, al ser la función inversa a la exponencial puede representar valores de entrada muy grandes en intervalos bastante más pequeños. Por esa razón se presentan las gráficas 4.4 y 4.5 donde se puede apreciar la densidad del logaritmo de los datos.

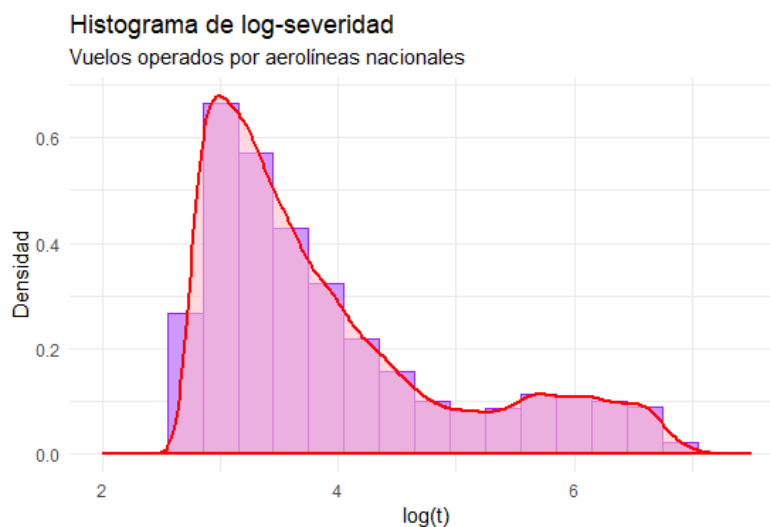


Figura 4.4: Log-severidad de aerolíneas nacionales

El aspecto más importante que se puede apreciar en la gráfica 4.4, es que se tienen dos módulos o campanas en la distribución. El primero es mayor al segundo y parece estar más sesgado positivamente mientras que el segundo que se encuentra alrededor del 6 en el eje  $X$ , parece estar más achatado. Esta observación, en principio no afecta la forma en que se realiza el ajuste, pero probará ser un indicador importante para cierto ajuste presentado más adelante.

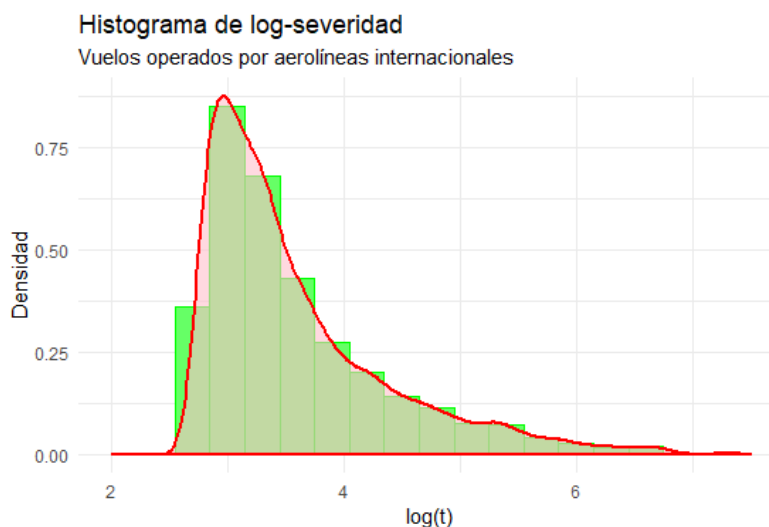


Figura 4.5: Log-severidad de aerolíneas internacionales

En la gráfica siguiente (4.5), no se puede observar existencia o al menos algún indicio de multimodalidad, sin embargo, obvia el sesgo positivo de la distribución a causa de valores lejanos a la mediana del vector.

Al igual que para los vectores de frecuencia, se presenta un listado (Tabla 4.4) con los valores de media, varianza y desviación estándar para los vectores de severidad.

	Agregados	Nacionales	Internacionales
Media	102.4596	107.7579	60.90952
Varianza	28944.61	30733.85	12968.52
SD	170.1312	175.3107	113.8794

Tabla 4.4: Valores de media, varianza y desviación estándar de severidad

Los valores presentados en la tabla son consistentes con las expectativas que se tenían, por ejemplo: se esperaba que el valor de la varianza de la severidad de incumplimiento de aerolíneas nacionales fuera bastante mayor que la de las internacionales, lo cual se puede observar de forma explícita, puesto que el valor de la desviación estándar del primer vector es aproximadamente 1.5 veces mayor que el segundo.



## 4.2. Ajuste del modelo agregado

Una vez hecho el análisis descriptivo de los datos, se presentará el proceso de modelado realizado. Así pues, en esta sección se describe la construcción del modelo agregado, mostrando primero el ajuste de distribución para la frecuencia y posteriormente a la severidad. Todo el trabajo de programación se realizó en el software estadístico **R**

Para el propósito de las pruebas que realizadas y que se presentan en éste capítulo, el nivel de significancia considerado es  $\alpha = 0.05$  y se siguió la premisa de que a mayor p-value, hay mayor evidencia para aceptar  $H_0$ : *Los datos siguen una distribución  $F_0$ .*

Para el ajuste de funciones de distribución típicas a los vectores se utilizó la función `goodfit()`<sup>2</sup> de la paquetería `vcd`, que esencialmente ajusta las distribuciones discretas a los datos en un vector mediante “*Maximum Likelihood Ratio Statistic*” o la prueba de  $\chi^2$  de Pearson. El ajuste se probó para las distribuciones Poisson, Binomial y Binomial Negativa y se presenta en las gráficas 4.6 y 4.7.

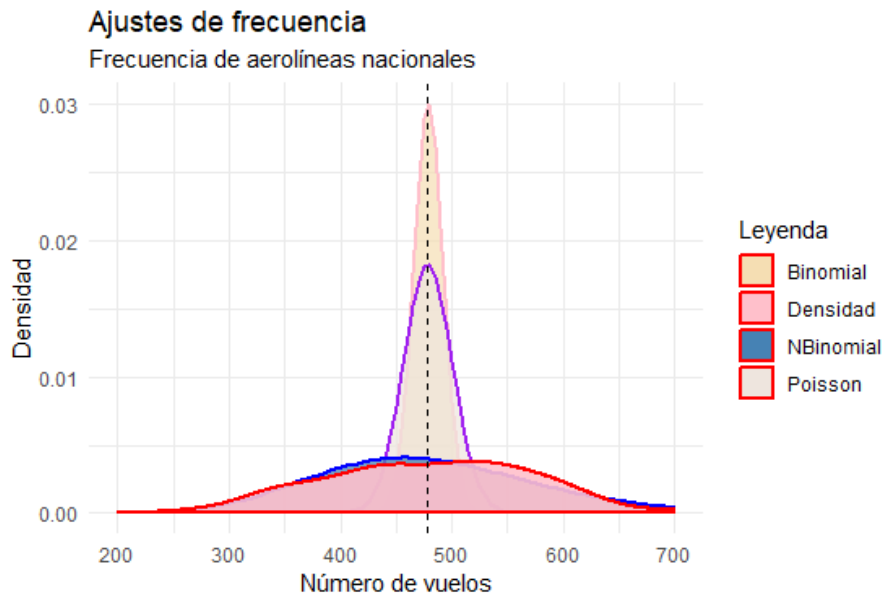


Figura 4.6: Ajustes a frecuencia nacional

<sup>2</sup>goodfit function, consulta 19/01/2018, fuente: <https://www.rdocumentation.org/packages/vcd/versions/1.4-4/topics/goodfit>

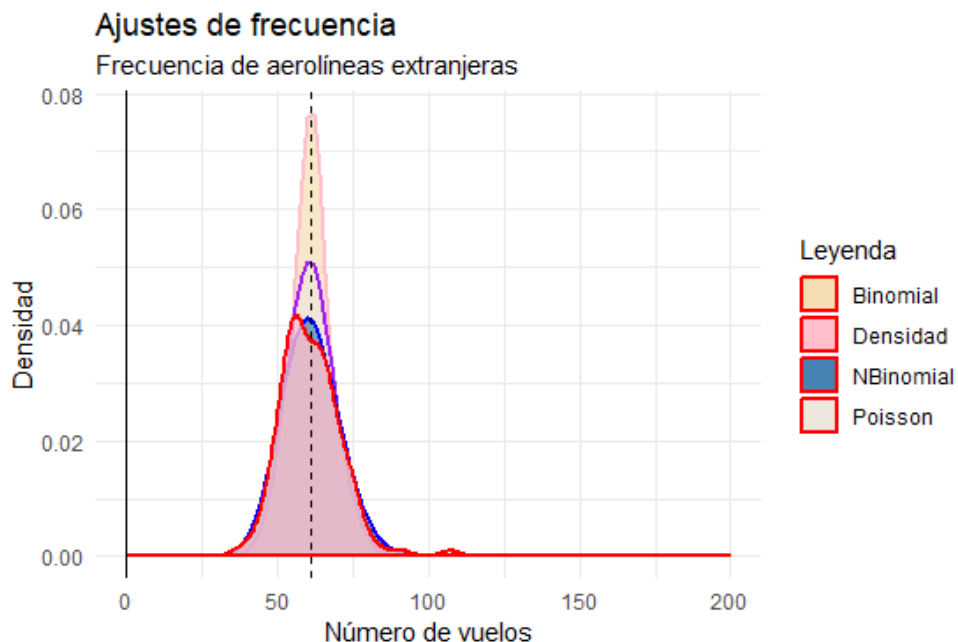


Figura 4.7: Ajustes a frecuencia internacional

Sobre esos mismos ajustes se realizaron las pruebas de bondad de ajuste Kolmogorov-Smirnov mediante la función  $ks.test()$ <sup>3</sup> y Anderson-Darling con la función  $ad.test()$ <sup>4</sup> en las tablas 4.5 y 4.6 se muestran los p-valores resultantes.

Distribución	P-value K-S	P-value A-D
Poisson	0	0
Binomial	0	0
Bin Negativa	0.4413	0.5156

Tabla 4.5: Resultados para frecuencia de aerolíneas nacionales

Distribución	P-value K-S	P-value A-D
Poisson	0.06524	0.135
Binomial	0	0
Bin Negativa	0.6816	0.6925

Tabla 4.6: Resultados para frecuencia de aerolíneas internacionales

<sup>3</sup> $ks.test$  function, consulta 18/05/2018, fuente: <https://www.rdocumentation.org/packages/dgof/versions/1.2/topics/ks.test>

<sup>4</sup> $ad.test$  function, consulta 18/05/2018, fuente: <https://www.rdocumentation.org/packages/nortest/versions/1.0-4/topics/ad.test>

Dado el nivel de confianza estipulado, se rechazan las hipótesis que la frecuencia de aerolíneas nacionales tiene distribución Poisson y Binomial, mientras que no existe evidencia suficiente para rechazar la hipótesis de que sigue una distribución Binomial Negativa.

En el caso de la frecuencia de aerolíneas internacionales, se tiene que tanto para la distribución Poisson como la Binomial Negativa, los valores de P-value obtenidos son mayores que el criterio de aceptación de 0.05. Sin embargo, considerando que los P-values obtenidos para la hipótesis de distribución Binomial Negativa son mayores en ambas pruebas, y para mantener mayor semejanza entre los modelos, se supondrá dicha distribución como la adecuada.

Por lo tanto,

$$\begin{aligned} N_{nac} &\sim BN(24.05119, 0.04784979) \\ N_{int} &\sim BN(108.9998, 0.6410929) \end{aligned} \tag{4.1}$$

Analizando las gráficas de severidad, se prevé el uso de distribuciones de colas pesadas como las mencionadas en el primer capítulo, sin embargo, para cubrir un rango más amplio de posibilidades, en ambos modelos se intentó realizar el ajuste tanto de distribuciones con colas más ligeras como la Gamma y la Exponencial, como distribuciones más adecuadas a modelos de riesgo con colas pesadas como lo son la Burr y la Pareto.

Dichos ajustes se basaron en la función *fitdistr()* de la paquetería *fitdistrplus*, la cual ajusta distribuciones univariadas a un vector de observaciones no censuradas,<sup>5</sup> mediante el método de máxima verosimilitud (elegido para los ajustes en esta tesis), método de momentos o método de cuantiles, como en las funciones de las pruebas de hipótesis Kolmogorov-Smirnov y Anderson-Darling.<sup>6</sup>

En la sección de anexos se incluye la función original de ajuste y aquí se presenta únicamente los resultados obtenidos de la misma.

Como se puede apreciar en las tablas 4.7 y 4.8, la pruebas de bondad de ajuste no arrojaron resultados que indiquen que alguna de las distribuciones ajustadas representen significativamente a los datos de severidad de las aerolíneas nacionales ni internacionales.<sup>7</sup>

---

<sup>5</sup>En el contexto estadístico, las *observaciones censuradas* representan observaciones parciales de la información. Este tipo de observaciones son comúnmente utilizadas en el estudio de la supervivencia, ya que las observaciones pueden representar ya sea fallos de los objetos de estudio o pérdida de información en un momento dado. Como estas observaciones tienen una interpretación diferente, su tratamiento en el cálculo de la verosimilitud también requiere un acercamiento diferente.

<sup>6</sup>*fitdistrplus* package, consulta 15/12/2018, fuente: <https://www.rdocumentation.org/packages/fitdistrplus/versions/1.0-11/topics/fitdist>

<sup>7</sup>Todos los valores de P-value obtenidos fueron menores a  $10^{-5}$  por lo que en la tabla se redondearon a 0. Los valores de los parámetros fueron redondeados a 5 decimales.

Distribución	Parámetro 1	Parámetro 2	P-value K-S	P-value A-D
Exponencial	0.00928	NA	0	0
Log-Normal	3.92773	1.07617	0	0
Gamma	0.7905235	0.0073366	0	0
Weibull Inversa	1.38525	31.84882	0	0
Gamma Inversa	1.43863	49.65623	0	0
Log-Logística	1.69604	43.04028	0	0

Tabla 4.7: Resultados para severidad de aerolíneas nacionales

Distribución	Parámetro 1	Parámetro 2	P-value K-S	P-value A-D
Exponencial	0.01642	NA	0	0
Log-Normal	3.60515	0.81268	0	0
Gamma	1.12878	0.018541	0	0
Weibull Inversa	1.92867	26.20730	0	0
Gamma Inversa	2.34722	68.75273	0	0
Log-Logística	2.37324	32.38397	0	0

Tabla 4.8: Resultados para severidad de aerolíneas internacionales

Lo que es más, para las distribuciones Burr, Pareto y Weibull, la función subyacente de optimización numérica de la verosimilitud, falló en encontrar valores óptimos de los parámetros que brindaran ajuste a la distribución empírica de los datos.<sup>8</sup>

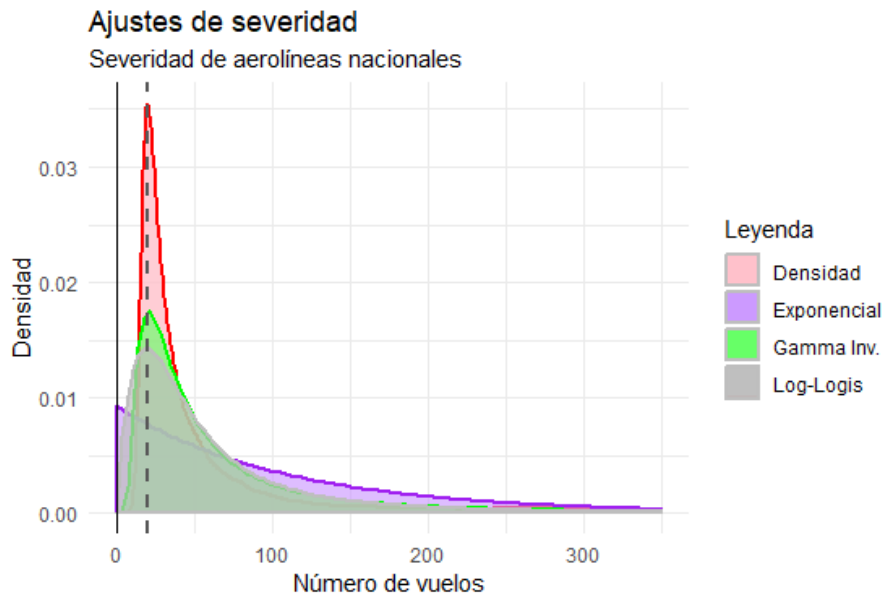


Figura 4.8: Ajustes a severidad nacional

<sup>8</sup>Estas distribuciones no se muestran en la tabla puesto que, al fallar el ajuste, no se pudo realizar la ejecución de las funciones *ks.test()* y *ad.test()*.

En las gráficas 4.8 y 4.9 se muestran las funciones de densidad ajustadas, en contraste con la función de densidad empírica obtenida de los datos de severidad.

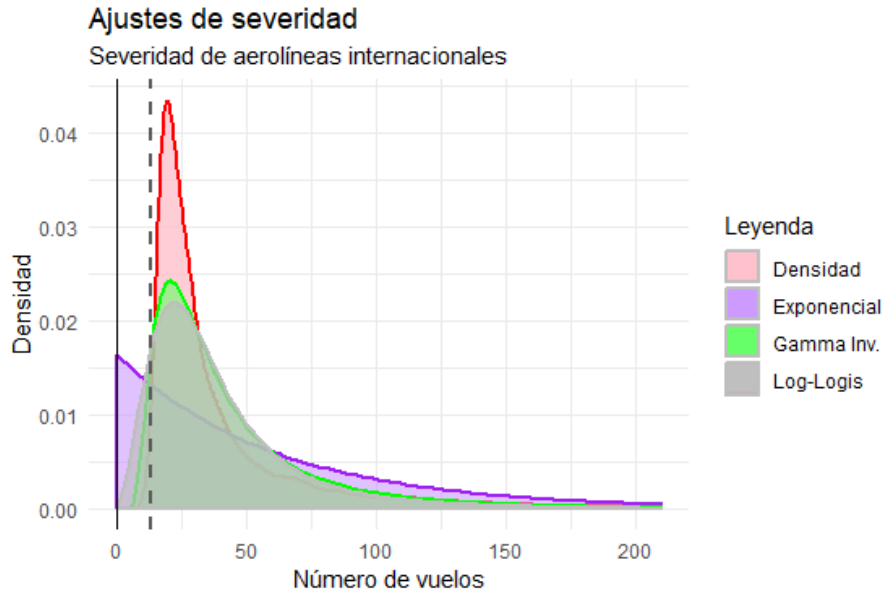


Figura 4.9: Ajustes a severidad internacional

Ante la problemática anterior, se decidió tomar un enfoque diferente e intentar modelar la severidad de los modelos mediante el uso de mezclas, ya que suelen ser más maleables y por lo tanto, brindar mayores posibilidades en situaciones como ésta en la que las opciones más comunes fallan.

Existen dos problemas a resolver en el ajuste de mezclas. El primero es elegir el número óptimo de componentes que conformarán la mezcla y el segundo es obtener el valor de los parámetros que maximicen la verosimilitud.

Para resolver el problema de número de componentes se tomaron dos enfoques, el primero es mediante la idea subyacente de que existe un número  $k^* \in \mathbb{N}$  tal que el aumentar el número de componentes a  $k^* + 1$  disminuirá el valor de la verosimilitud en lugar de incrementarlo. De manera práctica se traduce en calcular la verosimilitud para mezclas de 2, 3, 4, ... componentes y tomar las componentes para las que se obtenga un punto de inflexión en la gráfica de verosimilitud.

El segundo enfoque es suponer que existe una  $k^o \in \mathbb{N}$  (no necesariamente igual a  $k^*$ ), tal que la mezcla con  $k^o$  componentes proporcione un ajuste significativo bajo pruebas de bondad de ajuste para una  $\alpha$  dada.

Para facilitar la distinción de los enfoques se llamará al primero “enfoque por verosimilitud” y al segundo “enfoque por bondad de ajuste”.

Ya se ha presentado de forma teórica el algoritmo EM para encontrar el valor óptimo de  $\Theta$  para una mezcla de  $K$  distribuciones y que es la solución al segundo problema mencionado, sin embargo también formará parte de la estrategia que será utilizada para seleccionar el número de componentes óptimas en el modelo.

Si se aplica el algoritmo EM a ajustes para  $K = \{1, 2, \dots, n\}$  componentes, cada uno será el que maximice la función de verosimilitud dado un valor fijo  $K = k$  de componentes. Éstas verosimilitudes máximas serán las que se considerarán para la selección de componentes mediante ambos enfoques.

### Enfoque por verosimilitud

De manera general la función de verosimilitud está dada por  $L(\Theta) = \prod_{i=1}^n f(x_i, \Theta)$  y la Log-verosimilitud se puede ver como se muestra en el capítulo primero en la ecuación (2.6),

$$l(\Theta) = \sum_{i=1}^n \ln \sum_{k=1}^K \lambda_k f(x_i, \theta_k)$$

En particular para una mezcla de normales la ecuación sería:

$$l(\Theta) = \sum_{i=1}^n \ln \sum_{k=1}^K \lambda_k N(x_i, \mu_k, \sigma_k^2) \quad (4.2)$$

Para replicar la ecuación anterior se creó la función `loglik.norm.mix()` para generar el cálculo para una mezcla de  $k$  componentes. Para aplicar el enfoque se requiere aplicar la función para un conjunto más amplio de valores de  $k$ , por lo que se creó la función `verosimilitudes()` la cual genera un vector de log-verosimilitudes para mezclas de normales desde 1 hasta  $n$  componentes. Ambas funciones son presentadas en la sección de anexos.

En las gráficas (4.10) y (4.11), se presentan los valores obtenidos del segundo enfoque sobre los vectores de severidad para las dos clasificaciones de arribos.

En ambos casos el valor de log-verosimilitud que maximiza el algoritmo EM para mezclas de distribuciones normales se alcanza para mezclas con 9 componentes. En el caso de la severidad de incumplimiento para aerolíneas nacionales el máximo de log-verosimilitud es de  $-3718.303$ , pues para una componente más decrece a  $-3727.840$ . La misma situación ocurre para la severidad de las aerolíneas no nacionales donde el valor máximo es  $-4172.823$  y  $-4174.996$  para la mezcla con 10 componentes.

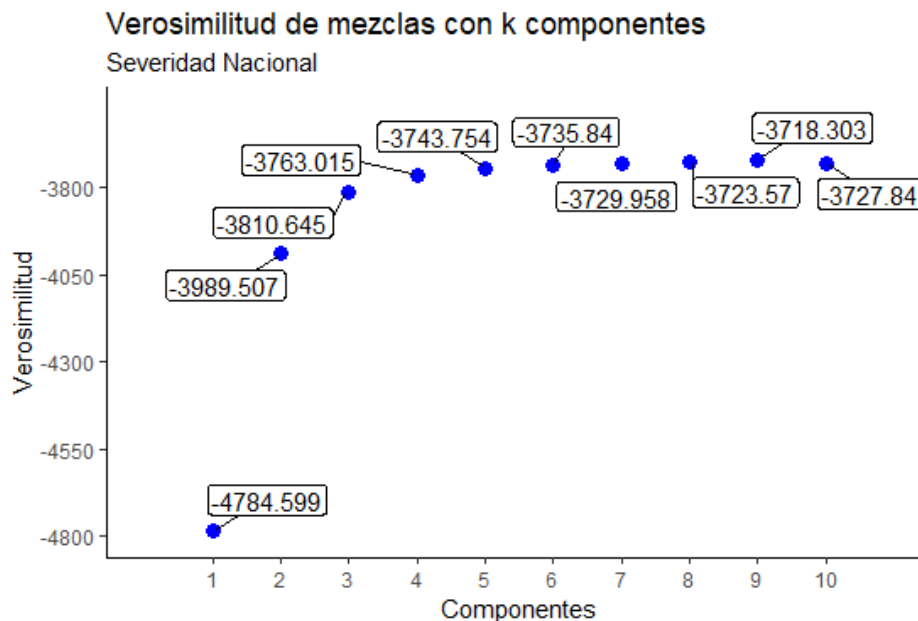


Figura 4.10: Log-verosimilitud para severidad de aerolíneas nacionales

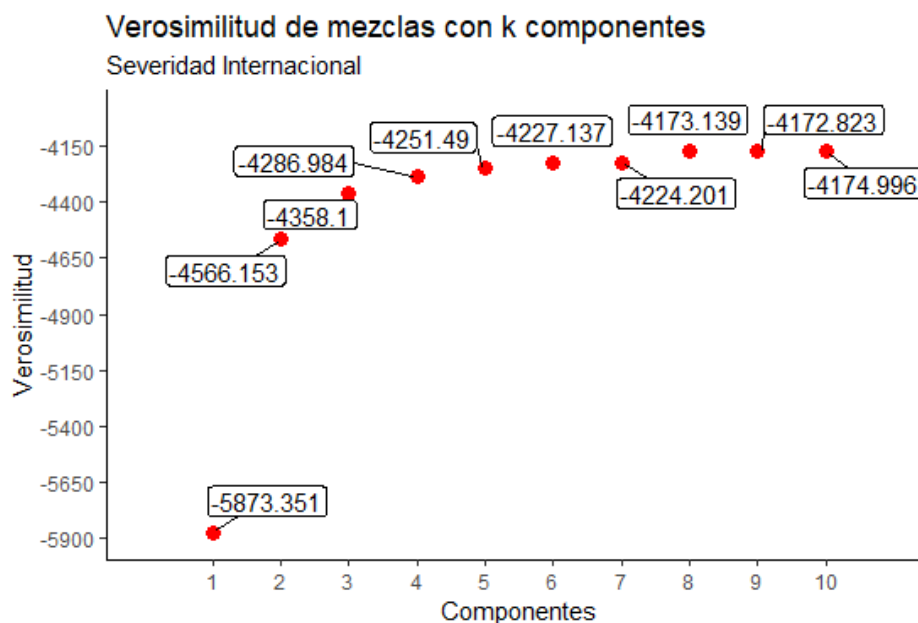


Figura 4.11: Log-verosimilitud para severidad de aerolíneas extranjeras

Las gráficas (4.12) y (4.13) muestran las diferentes componentes que conforman las mezclas que el enfoque sugiere como óptimas. Para cada distribución normal se calcularon los parámetros de media y desviación estándar, sin embargo en las gráficas que se presentan no se muestra como tal la densidad de 9 distribuciones normales con los parámetros estimados si no que se presentan considerando el peso que tiene cada una dentro del modelo.

La función  $normalmixEM()$  ordena las componentes con base en el valor de la media, lo cual busca enfrentar el problema de la identificabilidad mencionado previamente mediante el uso de lo que se conoce como “buenas prácticas” de la programación.

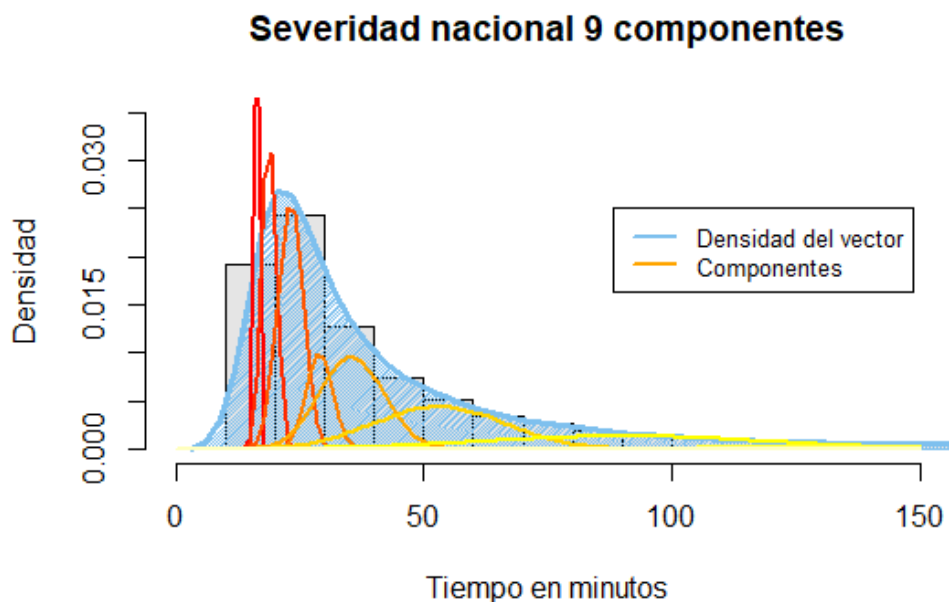


Figura 4.12: Componentes de severidad nacional

Una característica importante es el contraste entre los pesos de la segunda distribución y las dos distribuciones aledañas ya que, mientras que  $\lambda_2^9$  es la que tiene el peso mayor en la mezcla, con aproximadamente un quinto de la probabilidad (20.003%), los pesos  $\lambda_1$  y  $\lambda_3$  son los pesos más pequeños, 4.7% y 6.3% respectivamente.

Los valores de desviación estándar son crecientes al igual que la media con excepción de la segunda componente, teniendo esto en cuenta, en las primeras 7 componentes, que de hecho, representan poco más del 80% del peso total de la mezcla, se presentan incrementos relativamente pequeños tanto en media como en varianza lo cual sucede porque estas componentes están dando forma a la parte central de la densidad de la mezcla. Las últimas dos componentes de la mezcla tienen valores de media y varianza mucho más grandes que los de las componentes anteriores, y esto es porque en realidad conforman las colas de las distribución como se puede apreciar en la imagen (4.12).

Es decir, que en general se esperaría que aproximadamente el 80% de las observaciones generadas por el modelo provengan de las primeras 7 distribuciones confor-

<sup>9</sup>Las  $\lambda_j$  están dadas respecto a la media de las componentes siendo la de media menor  $\lambda_1$



mando el grueso de los valores típicos, mientras que una de cada cinco observaciones presenten valores más extremos en la cola de la distribución. Todo esto considerando que como la distribución normal está definida sobre la recta real pueden haber valores cercanos a la media de la mezcla que provengan de la dos últimas componentes.

La mezcla para la severidad de incumplimiento para vuelos operados por empresas internacionales tienen características similares a la mezcla para la severidad nacional, en cuanto a que, las primeras componentes conforman el cuerpo central de la distribución y son la octava y novena distribución las que dan forma y peso a la cola; sin embargo, este modelo tiene algunas características particulares que se presentarán a continuación.

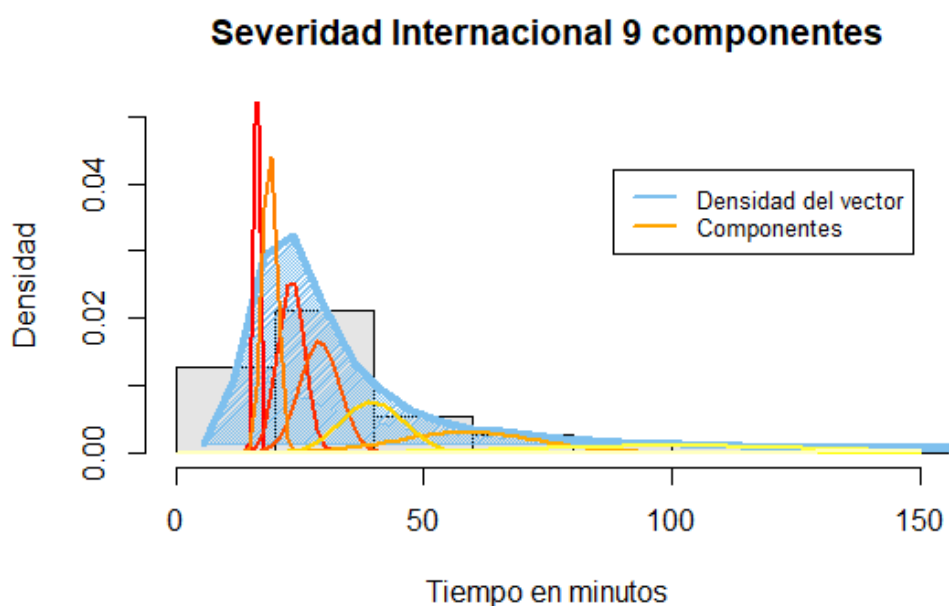


Figura 4.13: Componentes de severidad internacional

Los pesos de las primeras cuatro componentes son incrementales. Esto aunado a la representación gráfica del modelo conllevan a la suposición de que estas componentes no sólo forman parte de la campana de la densidad de la mezcla, si no que en particular forman la parte izquierda y central de la misma. Además, en la mezcla anterior el peso combinado de la octava y novena componente del modelo es cercano a 20 % mientras que en esta agregan apenas más que el 10 % generando un modelo donde las colas aportarán menos observaciones.

Curiosamente a pesar de que para este riesgo la media del modelo es menor, el rango en el cual se encuentran los valores de media y varianza es más amplio.<sup>10</sup>

<sup>10</sup>El valor de la media para el riesgo de incumplimiento en minutos de vuelos operados por aerolíneas nacionales es 1.769147 veces mayor al del riesgo por aerolíneas internacionales.

En ambos modelos todas las componentes tienen valores de media mayores a cero; además, sus valores de desviación (y por consiguiente de varianza) correspondientes, son pequeños para distribuciones con medias pequeñas y grandes sólo para aquellas con valores de media mayores. Esto a primera vista quita peso al problema de que pudieran generarse valores negativos entre las observaciones aunque naturalmente no lo erradica.

Está claro que si alguna de las componentes con valores de media más pequeños tuviera asociada una desviación grande, por ejemplo, si las primeras componentes con medias cercanas a 16.5 tuvieran los valores de desviación de la novena componente de su respectivo modelo, se esperaría que buena parte de las observaciones que generaran fueran negativas.

Para ver si la probabilidad de que el modelo genere observaciones menores a cero puede representar o no un problema o no para los modelos, se precisa calcular de manera explícita  $P[M < 0]$ . En ambos casos la probabilidad de obtener un valor negativo es menor al 0.1 %, por lo que se considera que no es un problema significativo para los modelos obtenidos.

### Enfoque por bondad de ajuste

La idea base de este enfoque es suponer que existe una  $k^o \in \mathbb{N}$  tal que la mezcla compuesta por  $k^o$  componentes proporcione un ajuste significativo bajo pruebas de bondad de ajuste para una  $\alpha$  dada. Es decir, que se buscará una mezcla que tenga menos componentes que la obtenida en el método anterior pero que mantenga la significancia en el ajuste que brinde.

Para ello, se utilizará la función `boot.comp()` de la paquetería `mixtools`<sup>11</sup> que realiza un proceso de bootstrap sobre los parámetros de la mezcla de manera secuencial para probar el número de componentes óptimas para modelos de mezclas.

En esencia, esta función realiza  $B$  muestreos con remplazo del vector  $x$  y utiliza estas muestras para hacer analíticamente la misma cantidad de pruebas de razón de verosimilitud para contrastar la bondad de ajuste entre una mezcla con  $k$  componentes y una con  $k + 1$  componentes, es decir,

$$\begin{aligned} H_0: & \text{ Los datos provienen de una mezcla } M \text{ con } k \text{ componentes} \\ H_1: & \text{ Los datos provienen de una mezcla } M' \text{ con } k + 1 \text{ componentes} \end{aligned}$$

Se realiza la misma ejecución de pruebas para tantos componentes como sea requerido y para cada una de ellas se calcula un p-value tal que, en el momento en el que sea mayor al valor de significancia para alguna componente, se pierde evidencia

---

<sup>11</sup>Package ‘mixtools’, consulta 06/12/2018, fuente: <https://cran.r-project.org/web/packages/mixtools/mixtools.pdf>

para rechazar la hipótesis nula y termina la ejecución de la función.

Por lo tanto, en este enfoque los modelos sugeridos tras la ejecución del código,<sup>12</sup> son mezclas de 3 y 4 componentes con distribución Normal para la severidad de incumplimiento en vuelos operados por aerolíneas nacionales y extranjeras respectivamente. A continuación, en la imagen 4.14 y en la tabla 4.9, se muestran las características y gráficas del modelo de 3 componentes para el primer vector.

Como se puede apreciar en la gráfica, la distancia entre la media de las tres componentes es relativamente amplia pero sigue la misma idea general que la mezcla de 9 componentes obtenida en el enfoque anterior. Es decir, las primeras componentes son las que describen la forma de la campana de la distribución y la última (o últimas en el enfoque por verosimilitud) describen los eventos extremos que conforman la cola de la distribución.

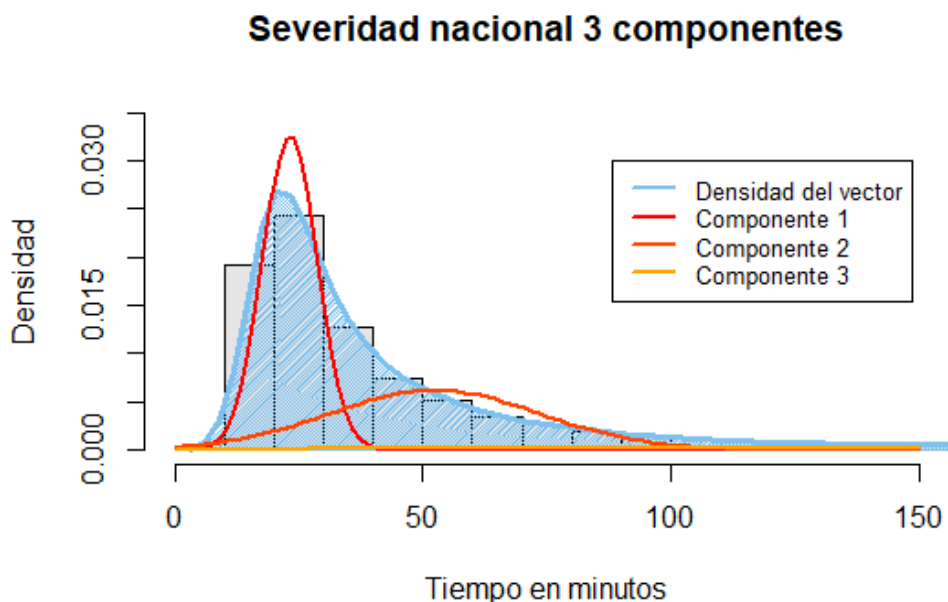


Figura 4.14: Mezcla de 3 componentes para severidad nacional

K	Lambda	Media	Desviación
1	0.4354035	23.20738	5.25919
2	0.3584901	59.67529	27.02119
3	0.2061065	385.87457	217.87322

Tabla 4.9: Parámetros para ajuste a severidad internacional (E2)

<sup>12</sup>La función y los resultados obtenidos se presentan en la sección de anexos.

Además de eso, en la misma gráfica se puede apreciar que las primeras dos componentes aparentan casi una bimodalidad respecto a la línea de la densidad empírica de los datos a diferencia de las gráficas de los dos modelos del enfoque anterior, donde los “bordes” de las densidades se acoplaban aproximadamente a la línea punteada que representa la densidad de sus respectivos vectores. Tomando la gráfica (4.4) donde se presenta la log-severidad de los vuelos operados, se aprecia que la forma es similar a la que se presenta bajo este enfoque ya que también se presentan dos módulos, donde el primero es el mayor en tamaño, lo cual es un indicio a que el modelo tiene buen ajuste.

A diferencia de los modelos correspondientes al enfoque anterior, en esta mezcla los pesos son decrecientes respecto al número de componentes y por tanto del valor de la media<sup>13</sup> lo cual naturalmente sugiere que hay más probabilidad de que una observación provenga de la primera componente que de la segunda y de la tercera, generando la campana al rededor de la mediana.

Además, el peso de la cola resulta consistente entre enfoques ya que en el primero se consideraba que la cola, consistente de dos distribuciones, acumulaba un peso de cerca del 20% de la densidad de la mezcla que es cercano al peso que el algoritmo EM asigna a la tercera componente de este enfoque.

En la gráfica 4.15 y la tabla 4.10, de manera similar a las anteriores, se presenta la información del modelo de 4 componentes para el ajuste de la severidad de los vuelos operados por aerolíneas extranjeras.

La diferencia más notoria entre enfoques para este caso, es que deja de empatar el número de componentes contra la mezcla recomendada para el riesgo de incumplimiento de operadoras aviarias nacionales; en este caso se pasó la prueba de hipótesis al contrastar el modelo que se presenta contra una mezcla con 5 componentes.

A pesar de esta diferencia, los valores mostrados en la tabla 4.10 reflejan un comportamiento similar al del enfoque anterior en cuanto a que, mientras los valores de media son mayores para el índice de componente  $k$  sea mayor, los pesos asignados son cada vez menores.

K	Lambda	Media	Desviación
1	0.3178252	20.88409	3.395242
2	0.2597083	34.36748	8.707772
3	0.2274671	75.80329	28.071735
4	0.1949993	401.14523	213.923408

Tabla 4.10: Parámetros para ajuste a severidad internacional (E2)

<sup>13</sup>Se había especificado que el ordenamiento de las componentes se da de menor a mayor valor de la media de la distribución por lo que para componentes con índices más grandes, la media es también mayor.

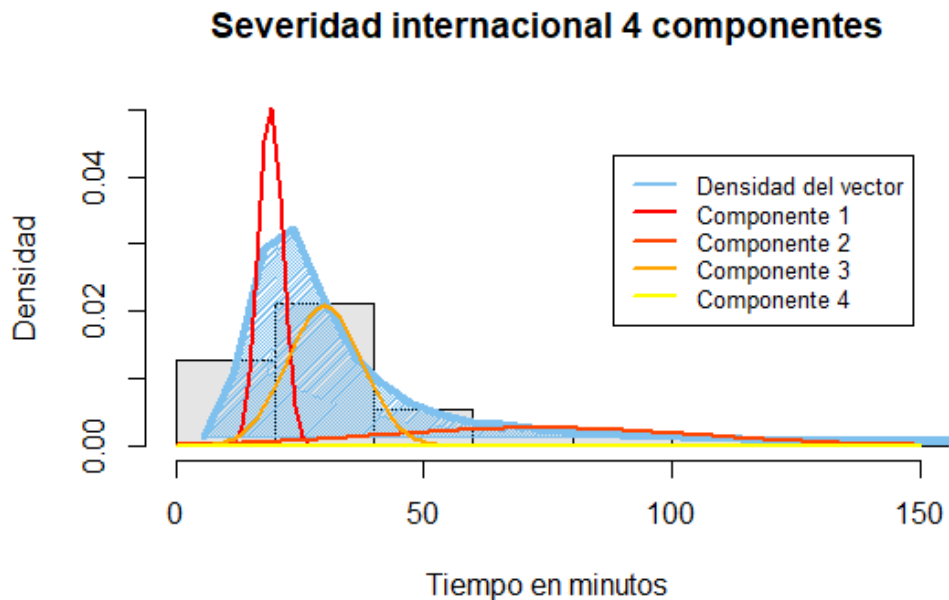


Figura 4.15: Mezcla de 4 componentes para severidad internacional

En esta ocasión se puede seleccionar<sup>14</sup> la cuarta componente como la que refleja el comportamiento de la cola del modelo; sin embargo, bajo esta selección el peso de la cola sería de aproximadamente 20% cuando en el enfoque anterior su peso era la mitad por lo que la comparabilidad no es muy buena.

### 4.3. Ajuste de aproximaciones

Existen varias aproximaciones al modelo agregado, el capítulo anterior se presentaron de manera teórica la aproximación Normal, la aproximación Log-normal y la aproximación Gamma trasladada por lo que a en esta sección se presentará el proceso práctico, mediante el cual, estas aproximaciones fueron aplicadas sobre los datos.

La aproximación Normal se basa en el Teorema Central del Límite, tal que dado un vector de datos, podemos ajustar una distribución normal  $N(\mu = \mathbb{E}[S], \sigma^2 = Var[S])$  siguiendo la idea que se desarrolló en la ecuación (3.4).

<sup>14</sup>Naturalmente esta selección de una componente que describe la cola de la mezcla es únicamente con fines ilustrativos para facilitar el análisis del modelo, bajo el entendimiento que un valor extremo tiene probabilidad de provenir de cualquiera de las componentes de la mezcla.

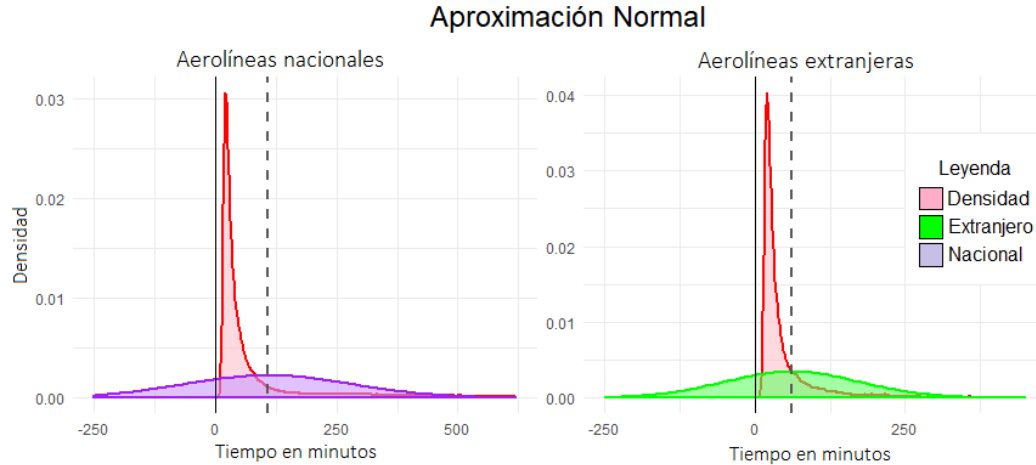


Figura 4.16: Aproximación Normal

En la gráfica izquierda se muestra la comparación de la densidad del vector de incumplimiento de aerolíneas nacionales con una distribución  $Normal(107.76, 175.31)$  y en la de la derecha la comparación para el de aerolíneas extranjeras con con una  $Normal(60.91, 113.87)$ . Se muestra de manera clara que los ajustes son muy pobres ya que difieren demasiado de las distribuciones empíricas.

En las gráficas se incluye una línea en color gris que marca la media de los valores observados, es decir, muestra el valor entorno al cual es simétrica la distribución de ajuste. Dados los valores de desviación estándar, vemos que en ambos casos la probabilidad de obtener un valor negativo es muy alta, de hecho es cercana a 0.5 por lo que no es un problema que pueda evitarse respaldándose en la regla empírica.

A continuación se presenta la aproximación Log-Normal. Usando como guía la teoría presentada en el capítulo anterior, se ajustó una distribución log-normal con parámetros de media ( $meanlog$ ) 3.927732 y desviación estándar ( $sdlog$ ) 1.076171 para el primer vector de severidad y parámetros 3.6051534 y 0.8126837 para el segundo.<sup>15</sup> El resultado de dichos ajustes se presentan en las gráficas comparativas en la imagen 4.17.

Como era de esperar, estas aproximaciones tienen un mejor ajuste gráfico a la distribución empírica distribuyendo de manera más apropiada las funciones de densidad ya que al no tener probabilidad positiva sobre los reales negativos y no ser simétricas, se recarga mayor probabilidad entorno a la moda.

A pesar de que se haya mejorado el modelo aproximado, es notoria la falta de exactitud al ajustar el peso de las colas en la distribución, ya que como se había mencionado con anterioridad, existen datos en la muestra cuyos valores son muy grandes.

<sup>15</sup>Los parámetros  $meanlog$  y  $sdlog$  son los valores de media y varianza de la distribución dados en escala logarítmica. Los valores default para la parametrización en  $\mathbf{R}$  son 0 y 1 respectivamente.

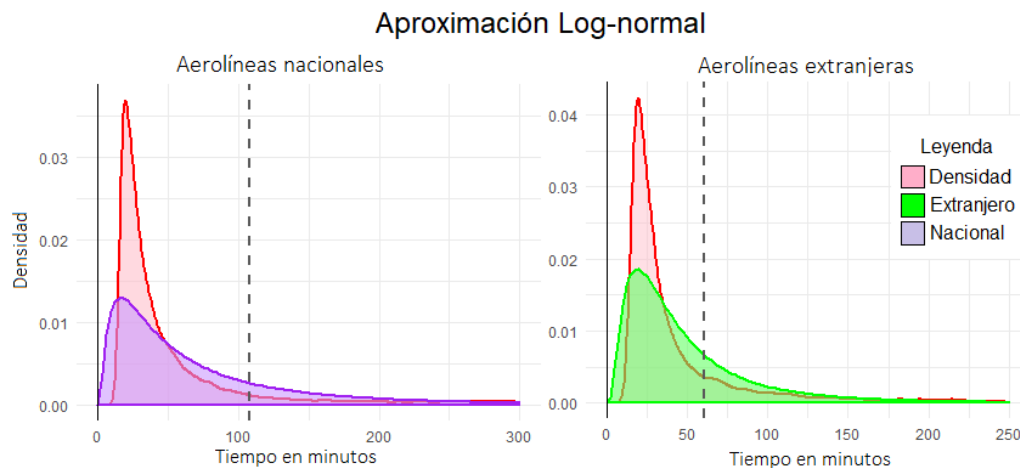


Figura 4.17: Aproximación Log-Normal

Aún queda por presentar el tercer ajuste del modelo, el ajuste por distribución gamma trasladada. Como se presentó en el capítulo anterior el ajuste de esta distribución se basa en el ajuste de una distribución gamma “común y corriente” sobre el vector trasladado  $X - x_0$ . Esto naturalmente afecta la distribución de probabilidad sobre el soporte.

Entre las características de los datos que se recompilaron de la página de internet del AICM, se mencionó que los registros en la base no son los registros naturales de los arribos sino que, siguiendo las normas aeroportuarias, contiene únicamente el registro de los vuelos que llegan con 15 o más minutos de retraso en la llegada. El ajuste de esta aproximación resultó en valores de  $X_0$  muy cercanos a 15 por lo que se decidió redondearlos a este número para conservar la regla utilizada en el aeropuerto.

En las gráficas de la aproximación (4.18 y 4.19) se muestra no sólo la densidad empírica de los vectores junto a la densidad teórica del ajuste, sino que se agregó la densidad de un ajuste Gamma no trasladado para poder apreciar las diferencias entre bondades de ajuste gráficas.

A pesar de que los dos ajustes a la distribución empírica presentados en esta primera gráfica son muy pobres se puede ver una gran diferencia entre el ajuste gamma y el gamma trasladada, pues en el primero la cola de la distribución es mucho más pesada que la real y la forma de la función para valores entre cero y la mediana es para nada similar. Ambos fallos en el ajuste mejoran un poco para la aproximación trasladada pues se aligera la cola y se acerca la curva del módulo de la aproximación a la empírica.

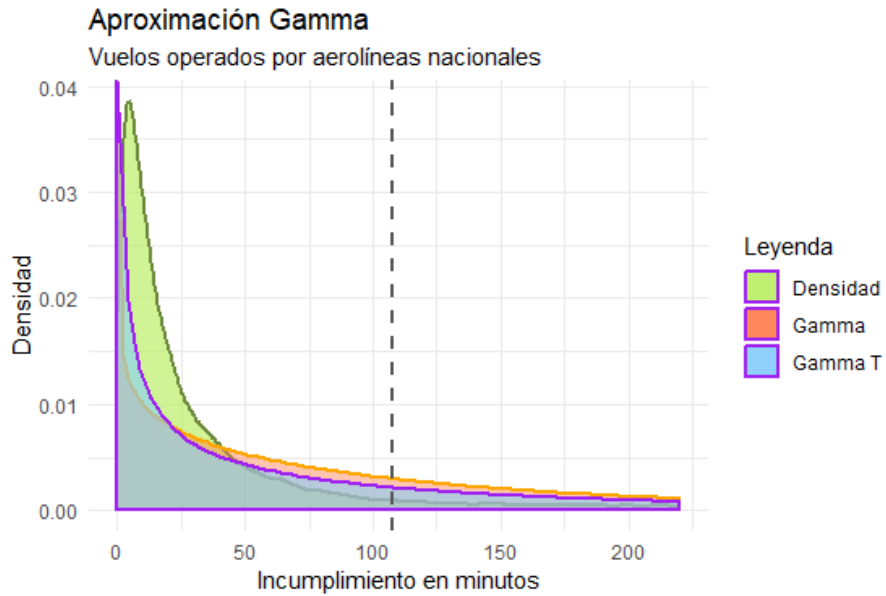


Figura 4.18: Aproximación Gamma trasladada

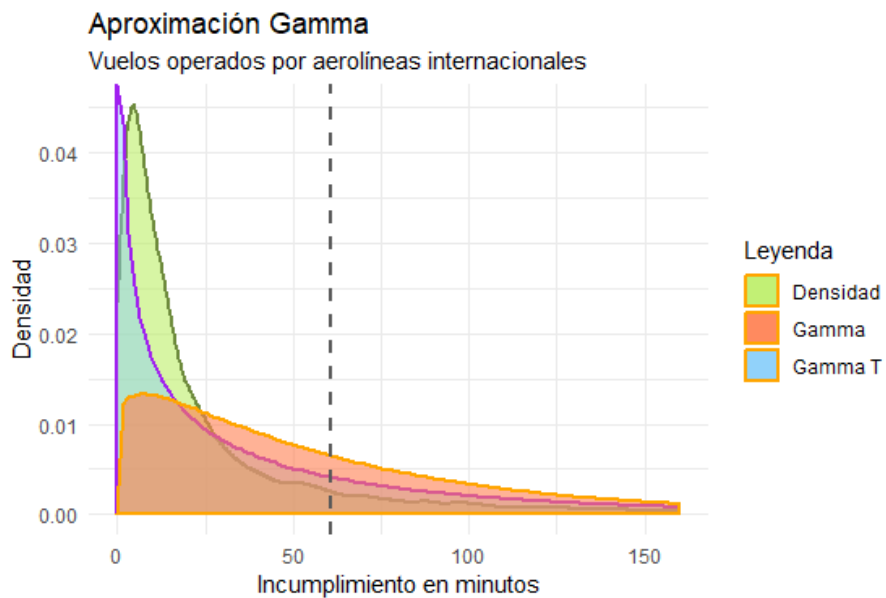


Figura 4.19: Aproximación Gamma trasladada

Similar a la primera gráfica, es notorio a simple vista tanto que el ajuste de la distribución gamma trasladada es mejor que el de la distribución gamma como el pobre ajuste general que brindan ambos modelos.



## 4.4. Interpretación y selección del modelo

La base del modelo agregado son las variables  $N$  y  $X$  que describen la frecuencia y severidad del riesgo por lo que, para tener la información del modelo, hay que tener primero información de ambas partes. En la primer parte de este capítulo se presentó el trabajo práctico realizado para obtener un ajuste significativo a cada una de estas partes así como diversas funciones auxiliares, estrategias de ajuste y más, pero no se concretó algún resultado para poder constituir el modelo agregado.

Respecto a las variables de frecuencia no hubo mayor problema para encontrar distribuciones que pasaran las pruebas de bondad de ajuste Kolmogorov-Smirnov y Anderson-Darling bajo un nivel de significancia  $\alpha = 0.05$ . Para describir las variables de frecuencia de riesgos se supusieron las distribuciones:

$$N_{nac} \sim BN(24.05119, 0.04784979).$$

$$N_{int} \sim BN(108.9998, 0.6410929).$$

Para las variables de severidad se intentó el mismo enfoque de ajuste sobre 9 distribuciones continuas pero la mayoría de estos fallaron como se presentó en las tablas 4.7 y 4.8, por lo que se optó por intentar ajustar una mezcla de distribuciones normales y proponer algunas aproximaciones.

En total se han presentado 5 modelos para describir la variable de severidad de cada riesgo: un primer ajuste de una mezcla de normales bajo el enfoque por verosimilitud, un segundo ajuste de mezcla de normales bajo el enfoque por bondad de ajuste, la aproximación Normal, la aproximación Log-normal y finalmente la aproximación Gamma trasladada.

Mediante los dos enfoques para el ajuste de mezclas se obtuvieron resultados bastante diferentes en pesos, número de componentes y sus parámetros, forma y más, pero para poder elegir una representación apropiada de la severidad debe preguntarse qué significan los resultados obtenidos. En esencia, al ajustar una mezcla de distribuciones, puede pasar una de dos cosas; la primera es que cada una de las  $k$  componentes de la mezcla representa un tipo diferente de vuelo, es decir que las características fundamentales de las observaciones procedentes de la componente  $j$ -ésima son en cierta medida diferentes a aquellas procedentes de la  $i$ -ésima y por ello, a pesar de pertenecer al mismo modelo y al mismo riesgo existen características que los diferencian como por ejemplo el modelo del avión utilizado en el vuelo, lugar de origen, ruta, clima, etc. Lo segundo que puede pasar es que las componentes sean simplemente la aproximación a una distribución fundamentalmente diferente.

A partir del primer enfoque se obtiene el mejor modelo de mezclas posible para el vector analizado, es decir que no existe otro modelo de mezclas de normales que aporte mayor información. Por lo tanto si los resultados del enfoque se interpretan

como tipos de vuelos, la forma para obtener la mayor información posible sería categorizando los vuelos en nueve grupos diferentes y únicas indicando los criterios de segregación entre categorías.

Sin embargo, en las gráficas podemos ver que en realidad los modelos propuestos no presentan multimodalidad o algún indicio de que la mejor implementación del modelo sea mediante la categorización de las muestras, por lo que conservar mezclas de 9 componentes para cada v.a. de severidad parece algo excesivo.

Para intentar describir la severidad haciendo uso de mezclas que consistieran de menos componentes, mientras que mantuvieran bondad de ajuste se realizó el modelado mediante el segundo enfoque, en el cual, la base de la selección es que de manera individual para cada vector de severidad, se comparase la mezcla de dos componentes con la de tres, la de tres con la de cuatro y así sucesivamente hasta que se fallara en rechazar la hipótesis nula a nivel de significancia  $\alpha = 0.05$  por lo que se da pie a la presunción de que los datos provienen de una mezcla con  $k$  componentes.

Del segundo enfoque se obtuvo que para la severidad nacional es apropiada una mezcla de tres componentes y para la internacional una de cuatro componentes. Esto puede considerarse como la solución directa del problema de parsimonia en el primer enfoque ya que aunque no es el mejor modelo en términos de verosimilitud, la diferencia de log-verosimilitud entre uno y otro modelo es pequeña mientras que disminuye bastante la cantidad de parámetros por estimar.

Con ánimos de buscar un ajuste que cumpliera en mayor medida con el principio de parsimonia fue que se realizaron las aproximaciones, sin embargo, las tres generaron ajustes pobres; la primera por la alta probabilidad de obtención de valores negativos, la segunda por el mal ajuste a la campana de la distribución y la tercera por la forma y ajuste general a la distribución empírica.

Por tanto, las mezclas obtenidas por el segundo enfoque, de tres y cuatro componentes normales, son las seleccionadas para ajustar a las variables de severidad dentro de los modelos agregados respectivos.

## 4.5. Modelos finales y su simulación

Una vez habiendo seleccionado cómo se modelará cada una de las partes por separado es posible obtener información precisa del modelo agregado. En esta sección se presentan métricas de las componentes teóricas y su agregación en la simulación de un periodo de actividad en el aeropuerto para observar la calidad del modelo como medida de análisis del riesgo sobre el tiempo.

### Modelo para el riesgo de incumplimiento para operadoras nacionales

El primer modelo se construyó con base en la información registrada para vuelos que fueron operados por aerolíneas mexicanas dentro del periodo de observación. En él se seleccionó la distribución Binomial Negativa para la frecuencia y una mezcla de 3 componentes normales para la severidad por lo que para cada variable se tiene:

*Frecuencia:*

- $\mathbb{E}[N] = \frac{kp}{1-p} \approx 478.5882$
- $Var[N] = \frac{kp}{(1-p)^2} \approx 10001.8875$

*Severidad:*

- $\mathbb{E}[M] = \sum_{i=1}^3 \lambda_i \mathbb{E}[X_i] \approx 107.7537$
- $Var[M] = \sum_{j=1}^k \lambda_j (\mu_j^2 + \sigma_j^2) - (\sum_{j=1}^k \lambda_j \mu_j)^2 \approx 30732.39$

Que es en extremo similar a los resultados empíricos del vector de datos donde la media y varianza tienen valores de 478.6 y 9,254 para la frecuencia, y 107.7579 30,733.85 para la severidad. Naturalmente existen pequeñas diferencias entre las medidas estadísticas empíricas y las obtenidas del modelo, sin embargo se debe tener en consideración que el modelo es una *aproximación* de un sistema altamente complejo.

### Modelo para el riesgo de incumplimiento para operadoras extranjeras

Para el modelo construido para describir el riesgo de incumplimiento en el tiempo de arribo en vuelos que fueron operados por aerolíneas extranjeras se seleccionó un modelo similar, donde la principal diferencia está en que la mezcla está conformada por 4 componentes y no 3. A continuación se muestran los resultados de este modelo teórico.

*Frecuencia:*

- $\mathbb{E}[N] = \frac{kp}{1-p} \approx 61.0222$
- $Var[N] = \frac{kp}{(1-p)^2} \approx 95.1846$

*Severidad:*

- $\mathbb{E}[M] = \sum_{i=1}^3 \lambda_i \mathbb{E}[X_i] \approx 60.8997$
- $Var[M] = \sum_{j=1}^k \lambda_j (\mu_j^2 + \sigma_j^2) - (\sum_{j=1}^k \lambda_j \mu_j)^2 \approx 12965.53$

Del mismo modo que en el modelo anterior, las diferencias entre los valores obtenidos del modelo y los obtenidos de los datos son muy pequeñas siendo apenas el resultado del proceso de ajuste. De esto se puede apreciar que la pérdida de información es muy poca en ambos casos, por lo que, además de generar buenos resultados, el proceso resulta bastante eficiente.

### Simulación y comparación de modelos

Como se comentó anteriormente, para poder observar la calidad del modelo como medida de análisis del riesgo sobre el tiempo, sigue probar la consistencia que tienen los modelos teóricos contra la realidad realizando la simulación de los riesgos y el análisis comparativo entre los datos históricos y la simulación.

Para poder simular el funcionamiento del aeropuerto en un intervalo de tiempo mensual, semestral, anual o similar, primero se debe poder simular una muestra de tamaño fijo. El tamaño de dicha muestra deberá ser determinado por una variable aleatoria con las características elegidas para la frecuencia del modelo y realizarse tantas veces como días por simular. Las funciones utilizadas para esto se presentan en la sección de anexos.

La función final fue ejecutada para cada uno de los modelos para simular 30 días de arribos al AICM considerando que corresponde a un mes bajo una convención de conteo 30/360. A continuación se presentan las gráficas de las distribuciones teóricas obtenidas para la construcción del modelo contra las gráficas de los vectores de severidad y frecuencia obtenidos de la simulación.

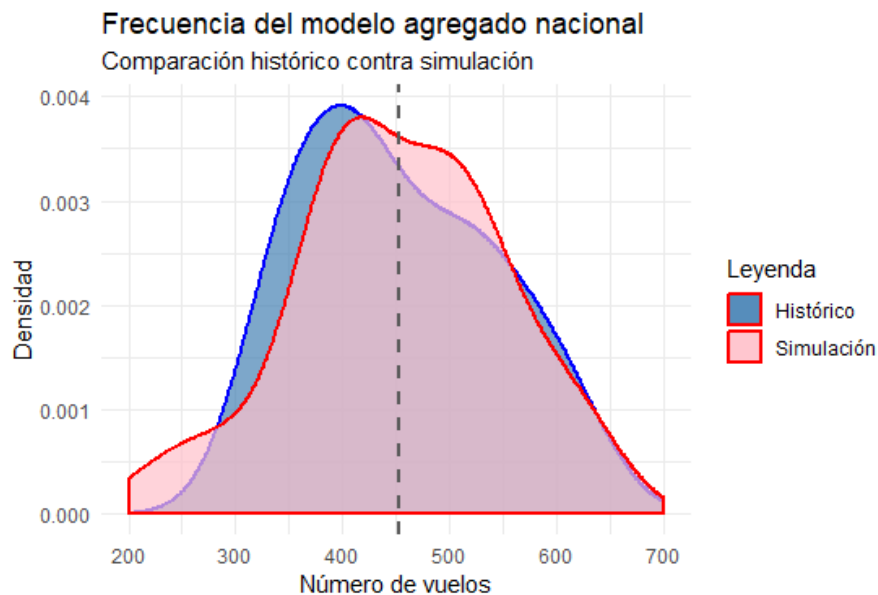


Figura 4.20: Comparativa resultados de frecuencia nacional

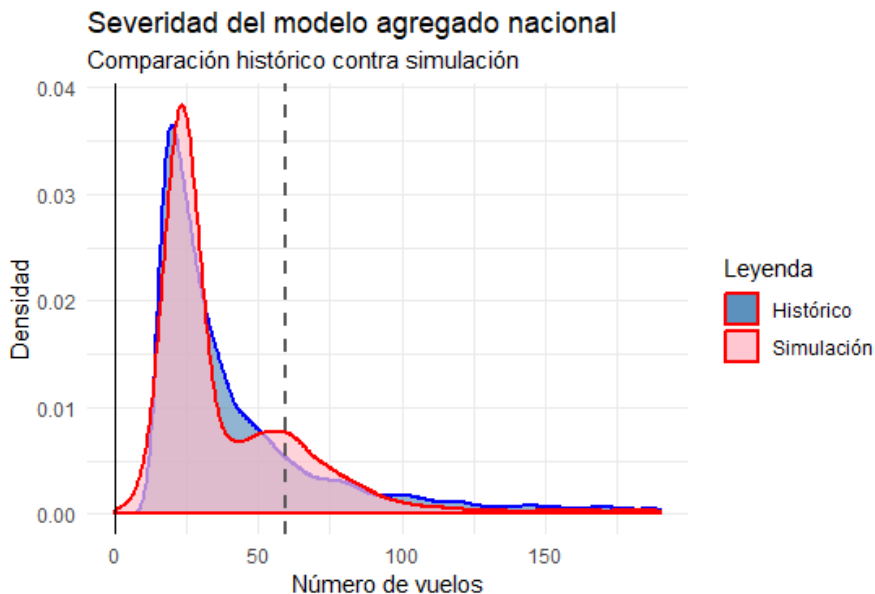


Figura 4.21: Comparativa resultados del modelo nacional

En las gráficas se divide el análisis en dos, en la gráfica 4.20 se presenta la densidad de la frecuencia simulada contra la frecuencia histórica marcando con una línea vertical la media del vector para poder apreciar el sesgo. En la gráfica 4.21, se presentan las distintas componentes de la mezcla y sobre él las densidades del vector muestra y del vector simulado.

En la gráfica de la variable de frecuencia (4.20) se puede ver el buen ajuste de la simulación de la variable, aunque naturalmente no es perfecto a causa de la aleatoriedad innata de la simulación. También conserva la forma en la campana de la función de densidad y un ligero sesgo.

De la simulación de severidad en la imagen (4.21) se puede notar cierta diferencia en la cola derecha pero este error es consistente con las componentes de la mezcla en severidad ya que la distribución tiene un segundo módulo que parece originarse de la segunda componente pero ligeramente ‘extendida’ a la derecha por los valores procedentes de la tercer componente que es la que genera los valores más extremos.

Considerando todo lo anterior se puede decir que el modelo construido es bueno para describir el riesgo de incumplimiento asociado a las aerolíneas mexicanas.

En la gráfica 4.22, se puede apreciar que la frecuencia simulada respecto al riesgo asociado a aerolíneas de origen extranjero, tiene una forma muy similar a los datos históricos en tanto que la severidad difiere en mayor medida.

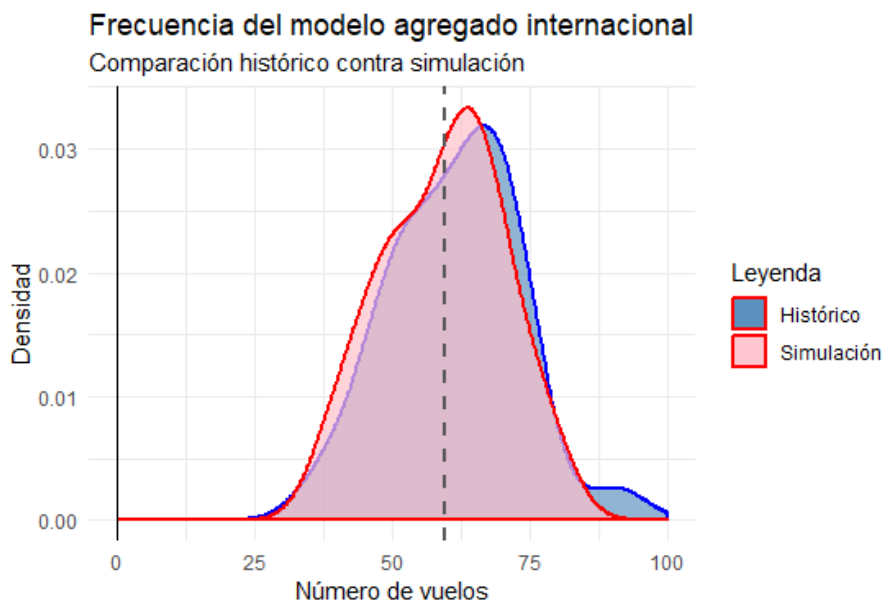


Figura 4.22: Comparativa resultados de frecuencia internacional

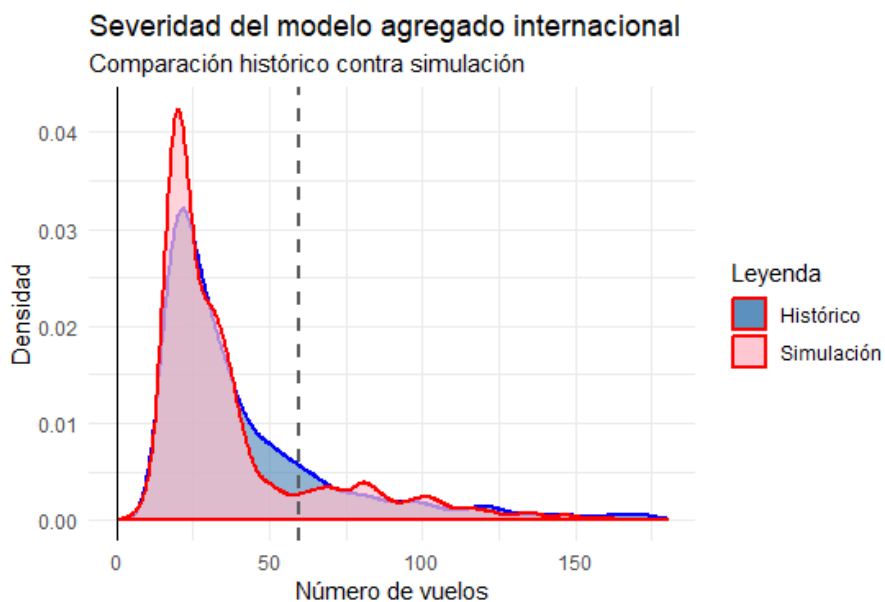


Figura 4.23: Comparativa resultados de severidad internacional

De la imagen 4.23 se observa que la simulación asigna mayor peso a valores al rededor de la mediana de los datos históricos a costo de los valores entorno al valor 50 por lo que la gráfica, en vez de mostrar la campana inclinada hacia la izquierda, presenta un valle contrastante con la densidad de los datos históricos. Sin embargo, hay que tener en cuenta que puede ser efecto de la realización de la clasificación aleatoria de las realizaciones dentro de una de las cuatro componentes, realizada por la función programada para la simulación.

Teniendo en mente la aleatoriedad como causante de las pequeñas diferencias entre el modelo teórico, los datos históricos y el vector simulado, se considerará que ambos modelos agregados seleccionados describen los riesgos de manera óptima.

# Conclusiones

La finalidad de este trabajo es establecer los conocimientos teóricos y prácticos para la construcción de dos modelos, uno para cada categorización del riesgo de incumplimiento en tiempo de arribo al aeropuerto de la Ciudad de México. Para ello, en el primer capítulo se presentaron distribuciones comúnmente utilizadas como ajustes a la distribución empírica de datos discretos y continuos, así como también se presentó el uso de pruebas de hipótesis para la selección de modelos.

Posteriormente, en el capítulo 2, se desarrolló la teoría de mezclas de distribuciones así como el proceso de ajuste de una mezcla mediante el uso del algoritmo EM desarrollado por Dempster, Laird y Rubin, con lo cual, se trata un sistema de ajuste que aunque conlleva más dificultades técnicas y procesamiento computacional, brinda enorme flexibilidad y maleabilidad de la forma de la distribución. De igual manera se presentan algunas problemáticas de convergencia y representabilidad en estos modelos.

En el capítulo siguiente se presentó el Modelo Compuesto o Agregado, en cierta medida, como ampliación del Modelo Individual, pues hace frente a algunos de los supuestos más restrictivos de este segundo modelo incorporando los conceptos de severidad y frecuencia. Así el modelo está dado por la agregación de una variable discreta de numeración de eventos y una variable continua que valora la magnitud del suceso.

En el cuarto y último capítulo, se unifica la teoría de distribuciones, ajustes, mezclas y aproximaciones dentro del Modelo Agregado, pues se realiza el proceso de ajuste sobre los datos de actividad del aeropuerto de la Ciudad de México. En este proceso se ajustaron funciones discretas comunes y ampliamente conocidas a la variable de frecuencia seleccionando una mediante la comparación de los resultados de las pruebas de bondad de ajuste; de manera similar, se intentó realizar el ajuste de distribuciones continuas con soporte no negativo para la severidad del modelo, sin embargo, no se obtuvieron los resultados deseados en el ajuste.

Como el ajuste de una única distribución no fue exitoso y considerando la amplia gama de factores que afectan las actividades aeroportuarias en distintos lugares del mundo, se decidió robustecer el proceso de modelado mediante la implementación del ajuste de mezclas y dos métodos para la valuación de su bondad de ajuste.



Así pues, se presenta el ajuste de mezclas de distribuciones Normales para cada riesgo, generando primero una mezcla que maximizara la Log-verosimilitud a costo de estimar una gran cantidad de parámetros, y posteriormente generando una segunda mezcla que buscara brindar no el mejor ajuste, sino un ajuste tan bueno como el realizador lo quiera, intentando mantener al mínimo la cantidad de parámetros estimados. A estos enfoques se les llamó en esta tesis, Enfoque por Verosimilitud y Enfoque por Bondad de Ajuste.

Finalmente, se presenta la selección de un modelo para cada clasificación del riesgo de incumplimiento en el tiempo de arribo al aeropuerto, seguida de las métricas y gráficas obtenidas de la simulación de los mismos, concluyendo así el proceso de modelado y validando su aplicabilidad.

A pesar de que en este trabajo se desarrolló el ajuste práctico para los datos de operación de un único aeropuerto, el autor considera que se cumplió el objetivo de no solo generar modelos para la descripción de los riesgos para un caso particular, si no asentar un proceso robusto en tanto a la teoría matemática como en el código informático que lo respaldan, que funja como el marco objetivo para fundamentar regulaciones en el sector aeroportuario, no solo a nivel local sino a nivel global.

# Bibliografía

- Andrew N. The EM algorithm. En *Lecture notes* (8). California, Estados Unidos: Stanford University.
- Charles M. Grinstead, J. Laurie Snell. (2006). Introduction to Probability. En *Distributions and Densities* (183-205). Estados Unidos: American Mathematical Society.
- Christian Kleiber, Samuel Kotz. (2003). Beta-Type Size Distributions. En *Statistical Size Distributions in Economics and Actuarial Sciences* (183-230). Estados Unidos: John Wiley and Sons, Inc.
- Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society, Series B.
- George Casella, Roger L. Berger. (2002). Hypotesis Testing. En *Statistical Inference* (423). California, Estados Unidos: Duxbury Advanced Series.
- Horst Rinne. (2008), Related distributions. En *The Weibull Distribution, A Handbook*. CRC Press.
- Horst Rinne. (2011). *A study of Weibull shape parameter: Properties and significance*. Reino Unido: Elsevier.
- Luis Rincón . Dos teoremas límite. En *Curso intermedio de probabilidad* (357-360). Ciudad de México: Facultad de Ciencias UNAM.
- Michael A. Stephens. (1979). *The Anderson-Darling Statistic*. California Estados Unidos: Stanford University.
- Murray R. Spiegel, Larry J. Stephens. (1989). The Standard Daviation and Other Measures of Dispersion. En *Schaum's Outline of Theory and Problems of Statistics* (95-101). Estados Unidos: McGraw-Hill.
- Prasanna Sahoo. (2013). Some Special Discrete Distributions. En *Probability and Mathematical Statistics* (107-129). Estados Unidos: University of Louisville.
- Sean Borman. (2004). *The Expectation Maximization Algorithm A short tutorial*. Utah, Estados Unidos: University of Utah, School of computing.

- Stuart A. Klugman, Harry H. Panjer, Gordon E. Willmot. (2004). *Loss Models: From Data to Decisions*. New Jersey, Estados Unidos: John Wiley & Sons.
- T. W. Anderson. (2010). The Anderson-Darling Statistic. En *Anderson-Darling Tests of Goodness-of-Fit* . California, Estados Unidos: Stanford University.
- Wayne L. Winston. (2004). Queuing Theory. En *Operations Research, Applications and Algorithms* (1051-1053). Estados Unidos: Thomson Learning, Inc.
- Wu, C. F. Jeff. (1983). On the Convergence Properties of the EM Algorithm. En *The Annals of Statistics*. Vol 11 (95-103). Estados Unidos: Institute of Mathematical Statistics.

# Anexos

## Función de ajuste de densidades para vectores de severidad:

La función presentada a continuación funciona con base en tres valores de entrada: el vector o base de datos, un vector con los nombres de las distribuciones para las cuales se quiere probar el ajuste y un entero que indica la cantidad de vectores dentro de la variable *datos*.

Para cada distribución en el vector de nombres se realiza el ajuste a los datos mediante la función *fitdist()* y el método de máxima verosimilitud para luego realizar las pruebas de bondad de ajuste Kolmogorov-Smirnov y Anderson-Darling. Así, al finalizar la ejecución de la función, para cada vector de datos se obtiene un *DataFrame* en el cual está registrado el nombre de las distribuciones probadas, así como los parámetros estimados para el ajuste, el valor de las estadísticas de prueba así como los P-valores obtenidos para cada una.

```
1 multidist<-function(datos , distvector ,n){
2 lista <- list()
3 for(i in 1:n){
4 DFrame<- data.frame("*" =
5   c("*", "*", "*", "*", "*", "*", "*", "*"))
6 if(class(datos)=="list" || class(datos) == "data.frame"){
7 vec<-as.numeric(as.character(unlist(datos[i])))
8 }else if(class(datos) == "numeric" || class(datos) == "integer"){
9   vec <- datos
10 }else if(class(datos) == "matrix"){
11   vec <- datos[,i]
12 }else{
13   print("Solo :numeric , matrix , integer , data.frame o list ")
14   break()
15 }
16 vec<-vec[!is.na(vec)]
17 nombres<-c(paste("Vector ",i), "Par1", "Par2", "Par3",
18 "Kolm-Smirn", "P-value", "Ander-Darl", "P-value")
19 KS<-matrix(nrow=2,ncol=9)
20 AD<-matrix(nrow=2,ncol=9)
```

```

21 estimated<-matrix(nrow=3,ncol=9)
22 DFrame<-data.frame(DFrame,"Informaci n"=nombres)
23
24 for(j in 1:length(distvector)){
25   if(distvector[j]=="Exponencial"){
26     fit.exp <- tryCatch(fitdist(vec,"exp",method="mle"),
27     error = function(e1){print(paste("En Exponencial",e1))})
28     if(exists("fit.exp")){
29       #Kolmogorov
30       s<-tryCatch(ks.test(vec,"pexp",
31       rate=fit.exp$estimate[[1]][[1]]),
32       error = function(e2){NA})
33       KS[1,j]<-tryCatch(s[[1]][[1]],error = function(e3){NA})
34       KS[2,j]<-tryCatch(s[[2]],error = function(e4){NA})
35       #Anderson
36       d<-tryCatch(ad.test(vec,pexp,rate=fit.exp$estimate[[1]]),
37       error = function(e5){NA})
38       AD[1,j]<-tryCatch(d[[1]][[1]],error = function(e6){NA})
39       AD[2,j]<-tryCatch(d[[2]],error = function(e7){NA})
40       #Parametros
41       estimated[1,j]<-tryCatch(fit.exp$estimate[[1]],
42       error=function(e8){NA})
43     }
44   }else if(distvector[j]=="Burr") {
45     fit.burr<-tryCatch(fitdist(vec,"burr",method="mle",
46     start=c(shape1=1,shape2=1,scale=1)),
47     error = function(e9){print(paste("En Burr:",e9))})
48     if(exists("fit.burr")){
49       #Kolmogorov
50       s<-tryCatch(ks.test(vec,"pburr",
51       shape1=fit.burr$estimate[[1]],
52       shape2=fit.burr$estimate[[2]],
53       scale=fit.burr$estimate[[3]]),
54       error = function(e10){NA})
55       KS[1,j]<-tryCatch(s[[1]][[1]],error = function(e11){NA})
56       KS[2,j]<-tryCatch(s[[2]],error = function(e12){NA})
57       #Anderson
58       d<-tryCatch(ad.test(vec,pburr,
59       shape1=fit.burr$estimate[[1]],
60       shape2=fit.burr$estimate[[2]],
61       scale=fit.burr$estimate[[3]]),
62       error = function(e){NA})
63       AD[1,j]<-tryCatch(d[[1]][[1]],error = function(e13){NA})
64       AD[2,j]<-tryCatch(d[[2]],error = function(e14){NA})
65       #Parametros

```

```

66   for(k in 1:3){
67     estimated[k,j]<-tryCatch(fit.burr$estimate[[k]],
68     error = function(e15){NA}))
69   }
70   } else if(distvector[j]=="Log-Normal") {
71     fit.lnorm<-tryCatch(fitdist(vec,"lnorm",method="mle"),
72     error=function(e16){print(paste("En Log-Normal:",e16))})
73     if(exists("fit.lnorm")){
74       #Kolmogorov
75       s<-tryCatch(ks.test(vec,"plnorm",
76       meanlog=fit.lnorm$estimate[[1]],
77       sdlog=fit.lnorm$estimate[[2]]), error = function(e17){NA})
78       KS[1,j]<-tryCatch(s[[1]][[1]], error = function(e18){NA})
79       KS[2,j]<-tryCatch(s[[2]], error = function(e19){NA})
80       #Anderson
81       d<-tryCatch(ad.test(vec,plnorm,
82       meanlog=fit.lnorm$estimate[[1]],
83       sdlog=fit.lnorm$estimate[[2]]), error = function(e20){NA})
84       AD[1,j]<-tryCatch(d[[1]][[1]], error = function(e21){NA})
85       AD[2,j]<-tryCatch(d[[2]], error = function(e22){NA})
86       #Parametros
87       for(k in 1:2){
88         estimated[k,j]<-tryCatch(fit.lnorm$estimate[[k]],
89         error = function(e23){NA})
90       }}
91     } else if(distvector[j]=="Pareto") {
92       fit.pareto<-tryCatch(fitdist(vec,"pareto",method="mle"),
93       error = function(e24){print(paste("En pareto:",e24))})
94       if(exists("fit.pareto")){
95         #Kolmogorov
96         s<-tryCatch(ks.test(vec,"ppareto",
97         shape=fit.pareto$estimate[[1]],
98         scale=fit.pareto$estimate[[2]]), error=function(e25){NA})
99         KS[1,j]<-tryCatch(s[[1]][[1]], error=function(e26){NA})
100        KS[2,j]<-tryCatch(s[[2]], error=function(e27){NA})
101        #Anderson
102        d<-tryCatch(ad.test(vec,ppareto,
103        shape=fit.pareto$estimate[[1]],
104        scale=fit.pareto$estimate[[2]]), error=function(e28){NA})
105        AD[1,j]<-tryCatch(d[[1]][[1]], error=function(e29){NA})
106        AD[2,j]<-tryCatch(d[[2]], error=function(e30){NA})
107        #Parametros
108        for(k in 1:2){
109          estimated[k,j]<-tryCatch(fit.burr$estimate[[k]],
110          error = function(e31){NA})

```

```

111   }}
112   } else if (distvector[j]=="Weibull"){
113     fit.weibull<-tryCatch( fitdist( vec , "weibull" ,method="mle" ) ,
114     error = function(e32){ print(paste( "En weibull:" ,e32))})
115     if(exists( " fit . weibull " )){
116       #Kolmogorov
117       s<-tryCatch(ks.test( vec , "pweibull" ,
118       shape=fit.weibull$estimate [[1]] ,
119       scale=fit.weibull$estimate [[2]] ) , error=function( e33 ){NA})
120       KS[1 , j ]<-tryCatch( s [[1]] [[1]] , error=function( e34 ){NA})
121       KS[2 , j ]<-tryCatch( s [[2]] , error=function( e35 ){NA})
122       #Anderson
123       d<-tryCatch(ad.test( vec , pweibull ,
124       shape=fit.weibull$estimate [[1]] ,
125       scale=fit.weibull$estimate [[2]] ) , error=function( e36 ){NA})
126       AD[1 , j ]<-tryCatch( d [[1]] [[1]] , error=function( e37 ){NA})
127       AD[2 , j ]<-tryCatch( d [[2]] , error=function( e38 ){NA})
128       #Parametros
129       for(k in 1:2){
130         estimated [k , j ]<-tryCatch( fit . burr$estimate [[k]] ,
131         error = function(e39){NA})
132       }}
133   } else if (distvector[j]=="Gamma"){
134     fit.gamma<-tryCatch( fitdist( vec , "gamma" ,method="mle" ) ,
135     error = function(e40){ print(paste( "En Gamma:" ,e40))})
136     if(exists( " fit . gamma " )){
137       #Kolmogorov
138       s<-tryCatch(ks.test( vec , "pgamma" ,
139       shape=fit.gamma$estimate [[1]] ,
140       rate=fit.gamma$estimate [[2]] ) ,
141       error = function(e41){NA})
142       KS[1 , j ]<-tryCatch( s [[1]] [[1]] , error = function(e42){NA})
143       KS[2 , j ]<-tryCatch( s [[2]] , error = function(e43){NA})
144       #Anderson
145       d<-tryCatch(ad.test( vec , pgamma ,
146       shape=fit.gamma$estimate [[1]] ,
147       rate=fit.gamma$estimate [[2]] ) ,
148       error = function(e44){NA})
149       AD[1 , j ]<-tryCatch( d [[1]] [[1]] , error = function(e45){NA})
150       AD[2 , j ]<-tryCatch( d [[2]] , error = function(e46){NA})
151       #Parametros
152       for(k in 1:2){
153         estimated [k , j ]<-tryCatch( fit . burr$estimate [[k]] ,
154         error = function(e47){NA})
155       }}

```

```

156 } else if(distvector[j]=="Gamma Inversa") {
157   fit.invgamma<-tryCatch(
158     fitdist(vec,"invgamma",method="mle"), error =
159     function(e48){print(paste("En gamma inversa:",e48))})
160   if(exists("fit.invgamma")){
161     #Kolmogorov
162     s<-tryCatch(ks.test(vec,"pinvgamma",
163       shape=fit.invgamma$estimate[[1]],
164       scale=fit.invgamma$estimate[[2]]), error=function(e49){NA})
165     KS[1,j]<-tryCatch(s[[1]][[1]], error=function(e50){NA})
166     KS[2,j]<-tryCatch(s[[2]], error=function(e51){NA})
167     #Anderson
168     d<-tryCatch(ad.test(vec,pinvgamma,
169       shape=fit.invgamma$estimate[[1]],
170       scale=fit.invgamma$estimate[[2]]), error=function(e52){NA})
171     AD[1,j]<-tryCatch(d[[1]][[1]], error=function(e53){NA})
172     AD[2,j]<-tryCatch(d[[2]], error=function(e54){NA})
173     #Parametros
174     for(k in 1:2){
175       estimated[k,j]<-tryCatch(fit.invgamma$estimate[[k]],
176       error = function(e55){NA})
177     }
178   }else if(distvector[j]=="Weibull Inversa") {
179     fit.invweibull<-tryCatch(
180       fitdist(vec,"invweibull",method="mle"), error =
181       function(e56){print(paste("En Weibull Inversa",e56))})
182     if(exists("fit.invweibull")){
183       #Kolmogorov
184       s<-tryCatch(ks.test(vec,"pinvweibull",
185         shape=fit.invweibull$estimate[[1]],
186         scale=fit.invweibull$estimate[[2]]),
187         error=function(e57){NA})
188       KS[1,j]<-tryCatch(s[[1]][[1]], error=function(e58){NA})
189       KS[2,j]<-tryCatch(s[[2]], error=function(e59){NA})
190       #Anderson
191       d<-tryCatch(ad.test(vec,pinvweibull,
192         shape=fit.invweibull$estimate[[1]],
193         scale=fit.invweibull$estimate[[2]]),
194         error = function(e60){NA})
195       AD[1,j]<-tryCatch(d[[1]][[1]], error = function(e61){NA})
196       AD[2,j]<-tryCatch(d[[2]], error = function(e62){NA})
197       #Parametros
198       for(k in 1:2){
199         estimated[k,j]<-tryCatch(fit.invweibull$estimate[[k]],
200         error = function(e63){NA})

```



```

201   }}
202   }else if(distvector[j]=="Log-Logistica") {
203   fit.llogis<-tryCatch(
204   fitdist(vec,"llogis",method="mle"), error =
205   function(e64){print(paste("En Log-log stica:",e64))})
206   if(exists("fit.llogis")){
207   #Kolmogorov
208   s<-tryCatch(ks.test(vec,"pllogis",
209   shape=fit.llogis$estimate[[1]],
210   scale=fit.llogis$estimate[[2]]),
211   error = function(e65){NA})
212   KS[1,j]<-tryCatch(s[[1]][[1]], error = function(e66){NA})
213   KS[2,j]<-tryCatch(s[[2]], error = function(e67){NA})
214   #Anderson
215   d<-tryCatch(ad.test(vec,pllogis,
216   shape=fit.llogis$estimate[[1]],
217   scale=fit.llogis$estimate[[2]]),
218   error = function(e68){NA})
219   AD[1,j]<-tryCatch(d[[1]][[1]], error = function(e69){NA})
220   AD[2,j]<-tryCatch(d[[2]], error = function(e70){NA})
221   #Parametros
222   for(k in 1:2){
223   estimated[k,j]<-tryCatch(fit.llogis$estimate[[k]],
224   error = function(e71){NA})
225   }}
226   } else {
227   print("ERROR: Distribucion erronea o desconocida")}
228   DFrame<-data.frame(DFrame,
229   "Distribucion" <- c(distvector[j],
230   estimated[,j], KS[,j], AD[,j]))
231   }##End for 1:j
232   DFrame <- as.matrix(DFrame)
233   DFrame1 <- DFrame[-1,-c(1,2)]
234   rownames(DFrame1) <- DFrame[-1,2]
235   colnames(DFrame1) <- DFrame[1,-c(1,2)]
236   lista[[i]] <- DFrame1
237
238   }##End for 1:n
239   names <- c()
240   for(i in 1:n) names <- c(names,paste0("Vector",i))
241   names(lista) <-names
242   return(lista)
243 }

```

### Funciones primer enfoque

A continuación se presentan las funciones que generan la estimación de log-verosimilitud empleada para la selección del número de componentes normales para el ajuste del enfoque por verosimilitud.

En la sección correspondiente del tercer capítulo, se presentó de manera breve el funcionamiento de la función *loglik.norm.mix()*

```

1 loglik.norm.mix<- function(x, mix) {
2   k <- length(mix$lambda)
3   dnorm.mix.j<- function(x, j) {
4     mix$lambda[j]*dnorm(x, mean = mix$mu[j],
5       sd = mix$sigma[j])
6   }
7   verosim <- sapply(1:k, dnorm.mix.j, x=x)
8   lk <- log(rowSums(verosim))
9   L<-sum(lk)
10  return(L)
11 }

```

La función interna *dnorm.mix()*, valúa la función de densidad de la *j*-ésima componente de la mezcla  $f(\cdot, \theta_j)$  sobre los valores *x* recibidos. Esta función se valúa para cada una de las *k* componentes sobre todos los valores *x* mediante la función *sapply()* generando la siguiente matriz:

$$\begin{bmatrix} \lambda_1 N(x_1, \theta_1) & \lambda_2 N(x_1, \theta_2) & \lambda_3 N(x_1, \theta_3) & \dots & \lambda_k N(x_1, \theta_k) \\ \lambda_1 N(x_2, \theta_1) & \lambda_2 N(x_2, \theta_2) & \lambda_3 N(x_2, \theta_3) & \dots & \lambda_k N(x_2, \theta_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_1 N(x_n, \theta_1) & \lambda_2 N(x_n, \theta_2) & \lambda_3 N(x_n, \theta_3) & \dots & \lambda_k N(x_n, \theta_k) \end{bmatrix}$$

Así la variable *lk* contiene un vector con longitud *n* definido por:

$$\begin{bmatrix} \log \sum_{k=1}^K \lambda_k N(x_1, \mu_k, \sigma_k^2) \\ \log \sum_{k=1}^K \lambda_k N(x_2, \mu_k, \sigma_k^2) \\ \vdots \\ \log \sum_{k=1}^K \lambda_k N(x_n, \mu_k, \sigma_k^2) \end{bmatrix}$$

Finalmente el valor que regresa la función es la suma del vector resultando en la log-verosimilitud como en la ecuación (4.2).

Ya que la función sólo calcula la verosimilitud para una *k* fija, para aplicar la lógica bajo la cual se guía el enfoque, se creó la función *verosimilitudes()* la cual emplea la función anterior para generar un vector en el cual se almacene el valor de

log-verosimilitudes para mezclas de normales desde 1 hasta  $n$  componentes. Dicha función se presenta a continuación:

```

1  verosimilitudes<-function(x,K){
2    Lmulti<-vector(mode="numeric",length=K)
3    for(k in 2:K){
4      mix<-normalmixEM(x, k = k, maxit = 1001,
5      epsilon=0.01,maxrestarts = 30)
6      Lmulti[k]<-loglik.norm.mix(x = x, mix = mix)
7    }
8    #k=1 se calcula aparte
9    Lmulti[1] <- sum(dnorm(x,mean(x),sd(x),log=T))
10   return(Lmulti)
11 }

```

### Función segundo enfoque

Los argumentos de la función *boot.comp()* utilizada para la selección del modelo mediante el segundo enfoque son: el vector de datos requeridos para el testing para modelos de mezclas, *mix.type* que especifica el tipo de análisis que se desea realizar, *sig* actúa como el nivel de significancia que proporcionará el criterio de aceptación para las pruebas de hipótesis, *B* es el número de realizaciones de Bootstrap para el cálculo de la estadística de razón de verosimilitud y finalmente *max.comp* que indica en número máximo de componentes para las cuales se realiza el proceso de testing.

A continuación se muestra la función sobre los vectores de severidad y los resultados obtenidos.<sup>16</sup>

```

1  > (nacional.boot<-boot.comp(X1,mix.type="normalmix",
2      sig=0.05,max.comp = 10,maxit=200,epsilon=.01))
3  Decision: Select 3 Component(s)
4  $'p.values'
5  [1] 0.00 0.01 0.12
6  $obs.log.lik
7  [1] 157979.440 24969.094 7682.862

```

<sup>16</sup>En el código se puede ver la inclusión de los argumentos de entrada *maxit* y *epsilon* que no se listaron dentro de los argumentos de la función *boot.comp()* ya que de hecho, son argumentos para la función *normalmixEM()* que actúa de manera subyacente.

```

1 > (nacional.boot<-boot.comp(X1,mix.type="normalmix",
2     sig=0.05,max.comp = 10,maxit=200,epsilon=.01))
3 Decision: Select 3 Component(s)
4 $'p.values'
5 [1] 0.00 0.01 0.12
6 $obs.log.lik
7 [1] 157979.440 24969.094 7682.862

```

### Probabilidad del soporte negativo de los modelos por verosimilitud

```

1 pnorm.mix<-function(x,mix){
2   p<-mix$lambda*pnorm(x,mean=mix$mu,sd=mix$sigma)
3   return(sum(p))
4 }
5 > pnorm.mix(0,mix.nac.9norm)
6 [1] 0.0008384464
7 > pnorm.mix(0,mix.int.9norm)
8 [1] 0.000583845

```

### Simulación y comparación de modelos

Como desarrollado en el cuerpo del trabajo, la generación de un muestra simulada tiene base en diferentes funciones. A continuación se presenta la primera de ellas para la *composición* de un vector de  $n$  muestras de una mezcla de normales.

```

1 composicion <-function(n, Mu, Sig, P){
2   observaciones <-numeric(n)
3   if(length(Mu)!=length(Sig))
4     {stop("Mu debe tener la misma longitud que sig")}
5   if(sum(P)!=1) {stop("P debe sumar 1")}
6
7   for(i in 1:n){
8     Y <- sample(length(Mu), size = 1, prob = P)
9     observaciones[i] <-rnorm(1, Mu[Y], Sig[Y])
10  }
11  return(observaciones)
12 }

```

La segunda función desarrollada en la sección de simulación fue creada para generar datos simulados de  $D$  días de actividad en el AICM considerando la estructura de los modelos seleccionados. Para cada uno de estos días el número de vuelos a simular

depende del valor tomado por la variable de frecuencia  $N$  que, como se había mencionado, tiene distribución Binomial Negativa en ambos casos, por lo que el input a demás incluir el número de días a simular y la información de la severidad mediante un objeto clase *mixEM*, debe incorporar un objeto clase *goodfit* con la información respectiva a la frecuencia.

Así pues, se precisa almacenar en una variable, el valor de la simulación de la frecuencia de incumplimientos en un día y luego ese valor es utilizado para generar la severidad de los incumplimientos mediante el uso de la función anterior.

```
1 modelo.sim <-function(D = 1, frec , sev){
2   N <-vector(mode="numeric" , length = D)
3   X <-c()
4   #para cada uno de los D dias que se van a simular
5   for(i in 1:D){
6     n <-rbinom(1, size=frec$par[[1]] ,
7       prob=frec$par[[2]])
8     X <-c(X,composicion(n, Mu = sev$mu,
9       Sig = sev$sigma , P = sev$lambda))
10    N[i] <-n
11  }
12  sim <-list(Dias = D, Frecuencia = N, Severidad = X)
13  return(sim)
14 }
```