3+

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

OBSERVACIONES ABERRANTES: UNA PERSPECTIVA BAYESIANA

T E S I S

PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A

OLGA PATRICIA MIJANGOS HUESCA

DIRECTOR DE TESIS

DR. GUSTAVO JAVIER VALENCIA RAMIREZ

CD. UNIVERSITARIA, D. F. OCTUBRE, 1990

INDICE

CAPITULO 1. INTRODUCCION

CAPITULO 2. METODOS BAYESIANOS

CAPITULO 3. FAMILIA EXPONENCIAL

CAPITULO 4. UN METODO BAYESIANO

CAPITULO 5. EJEMPLOS

APENDICES

CAPITULO 1

INTRODUCCION

1.1 Observaciones Aberrantes

El estudio de las observaciones aberrantes ha llamado mucho la atención en las dos últimas décadas [1]. Esto se hace evidente por la extensa literatura acerca del tema.

Existe una gran variedad de nombres para identificar a dichas observaciones. En español se les conoce como observaciones aberrantes, observaciones discordantes, observaciones anormales, etc.; en inglés reciben los nombres de 'outlier', 'contaminants', 'wilds', 'dirty data', 'sports', 'surprising values', etc. Por mencionar solamente algunos de los tantos nombres que se les han dado.

Del mismo modo, existen diversas definiciones de lo que es una observación aberrante. Hay autores que hacen distinción entre los nombres y otros que no. Como ejemplos tenemos:

Grubbs (1969) define:

> Un 'outlier' es una observación que está marcadamente desviada de los otros miembros de la muestra en que se encuentra.

Barnett y Lewis (1978) usan el término de 'outlier' pero mencionan que sería igual llamarlas discordantes o aberrantes. La definición que proponen es:

> Un 'outlier' es una observación (o subconjunto de observaciones) que parecen ser inconsistentes con el resto del conjunto de datos.

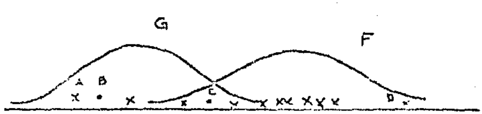
Box (1980) utiliza los nombres de 'outlier' y 'bad value' de igual manera.

[1] Balasooriya (1989).

Beckman y Cook (1983) mencionan cerca de siete nombres para dichas observaciones, pero sólo hacen distinción entre 3, dando las siguientes definiciones:

> Observación discordante.- Es cualquier observación que le es inesperada o le parece discrepante al investigador.
> Contaminante.- Es cualquier observación que no proviene de la distribución inicial asumida para la muestra.
> 'Outlier'.- En conjunto se refiere a cualquiera de las dos, contaminante u observación discordante.

Barnett (1983) comenta que el estudio de las observaciones aberrantes es un área estadística en la cual existe mucha confusión en las definiciones de los conceptos básicos y siente necesario que haya alguna estandarización. Apoya sus definiciones de acuerdo a varias necesidades, mencionando que al final, serán sólo su propio punto de vista. Acepta la definición de que un 'outlier' es una observación aberrante (o subconjunto de observaciones aberrantes) las cuales le parecen inconsistentes o discrepantes al investigador. Menciona que el término 'inconsistente' implica al menos de una manera informal el uso de un modelo inicial con respecto al cuál es inconsistente. En este sentido un 'outlier' es un valor extremo (o simplemente extremo), pero no necesariamente un extremo es un 'outlier'. Si el 'outlier' es estadísticamente irrazonable cuando se ha fijado como un extremo bajo el modelo inicial F, es un 'outlier' discordante. En un contexto de estudio de 'outliers', la muestra puede contener (pocas) observaciones que no provengan de F, sino de G, la cual pudiera ser un deslizamiento de F, o alguna distribución completamente distinta de F. El deslizamiento puede ser en la localización o en la dispersión, pero G necesita tener tal forma que las observaciones de G puedan manifestarse como extremos de la muestra (de este modo G no puede tener menos dispersión que F con una localización común). El número de observaciones de G en la muestra puede ser aleatorio o un número específico que sea pequeño en relación al tamaño de muestra. A tales miembros de G, Barnett les llama contaminantes u observaciones discordantes, aunque prefiere el primero pues el término de observación discordante se puede confundir con el de 'outlier' discordante. Aclara que un contaminante puede mostrarse como un 'outlier' o no y a su vez un 'outlier' puede ser un contaminante o no. Da un resumen de su explicación por medio del siguiente diagrama, donde × denota una observación de F y • una observacion de G.



A.. Es un 'outlier' que no es un contaminante. Posiblemente sea un 'outlier' discordante.

B. Es un contaminante y posiblemente un 'outlier' discordante.

C. Un contaminante que no es 'outlier'.

D. Un extremo que no es 'outlier'.

Balasooriya (1989) usa los términos de 'outlier', 'unusual observations' y 'spurious observations' indistintamente.

En el presente trabajo se hará distinción solamente entre observación aberrante y 'outlier'. El primer término englobará todos los nombres y definiciones anteriores y por 'outlier' denotaremos a aquellas observaciones aberrantes que sean extremos, pues como menciona Barnett (1983) puede haber observaciones aberrantes que no sean extremos.

1.2 Panorama Histórico

Al igual que existen distintos nombres y distintas definiciones acerca de las observaciones aberrantes, a través del tiempo se han tomado distintas actitudes hacia ellas. Actitudes para su detección y tratamiento.

Las primeras discusiones fueron hechas por Bernoulli (1777) quien menciona que no se debe rechazar una observación sólo por ser grande. Sus comentarios indican que la práctica de deshechar observaciones aberrantes éra muy común durante los 200 años anteriores [2].

Bessel (1848) argumenta que una observación no puede ser rechazada sólo por su valor grande y que necesariamente todas las observaciones deben ser incluidas con el mismo peso a contribuir en el resultado.

El primer criterio propuesto para rechazar observaciones aberrantes fué dado por el astrónomo Peirce (1852). Juzgando algunos comentarios de Peirce, nos damos cuenta que el rechazar observaciones aberrantes fué todavía muy común 75 años después de Bernoulli [3].

Tres años más tarde Gould (1855) aporta unas tablas como ayuda en el uso del criterio de Peirce.

Chauvenet (1863) reprodujo el criterio de Peirce pero basado en un argumento más simple. Su razonamiento fué el siguiente: Si $\theta(x)$ denota la probabilidad de que un error sea mayor igual que x,

> entonces el número de errores mayor o igual que x que pueden favorablemente esperarse en n observaciones es $n\theta(x)$. Por lo tanto, si encontramos x tal que $n\theta(x) = 1/2$, cualquier error más grande que x tendrá mayor probabilidad en su contra que a favor de él, y puede, por lo tanto, ser rechazado.

[2] Beckman y Cook (1983).

[3] Beckman y Cook (1983).

Otra propuesta para un rechazo cabal de observaciones aberrantes fué hecha por Stone (1868). Encontró que el uso del criterio de Peirce éra problemático y expresó dudas acerca de la exactitud de las matemáticas usadas. Stone también rechazó el método de Chauvenet ya que lo encontraba basado en un 'principio erróneo'. Razonó que Chauvenet al descartar una observación en 2 conjuntos de datos iguales, no tomaba en cuenta "el número promedio de observaciones que una persona toma con error", al cual llamo el 'módulo de descuido'. Es decir, un observador dado en una situación muestral dada comete en promedio un error en m observaciones tomadas. Se rechazarán todas las observaciones cuya probabilidad de ocurrencia sea menor que el módulo de descuido. La prueba es esencialmente similar a la de Chauvenet siendo idénticas si $m = 2n$.

La aparición de reglas de rechazo no fué muy bien recibida, pues se argumentaba que todas las observaciones deberían ser incluídas en el análisis.

Glaisher (1873,1874) fué tal vez el primero en publicar un procedimiento de ponderación, mencionando al respecto 'parece ser que reemplazará la necesidad de rechazar las observaciones aberrantes'. También fué el primero en sugerir que las observaciones pueden provenir de problaciones diferentes.

El mismo método fué publicado por Edgeworth (1883) 10 años mas tarde.

Otros métodos basados en ponderaciones fueron hechos por Newcomb (1886).

Edgeworth creo tres modelos para 'outliers' usando métodos de Monte Carlo. Los utilizó en un estudio para comparar los métodos de Glaisher, Newcomb, Peirce y Stone.

Encontró que el módulo de descuido de Stone éra legítimo. Reconoció los diferentes modelos bajo los trabajos de Stone y Glaisher. Indicó que el criterio de Chauvenet y Peirce eran muy pesimistas. Concluyendo así, que Newcomb estaba por encima de los otros.

Regresando a la historia de las pruebas de rechazo 'ad hoc' tales como las de Chauvenet y Stone, tenemos a Wright (1884) que propuso un procedimiento basado en el rechazo de cualquier observación que diste de la media por más de tres veces la desviación estándar.

Goodwin (1913) propuso otro procedimiento en el que se rechaza una observación aberrante en una muestra de tamaño n si ésta dista de la media de las $n-1$ observaciones restantes cuatro veces la desviación promedio de estas $n-1$ observaciones.

La investigación continúa hasta nuestros días desarrollándose muy variados y sofisticados métodos estadísticos para su detección y tratamiento.

1.3 Causas de los 'outliers'

Al realizar el estudio de un fenómeno a través de un conjunto de observaciones se verá que éstas varían. La variabilidad o dispersión en un conjunto de observaciones puede provenir de muy diferentes fuentes.

Tres fuentes de variabilidad en los registros son:

(i) Variabilidad inherente. Esta variabilidad no puede ser controlada, es propia del fenómeno.

(ii) Error de medición. Es el error que se comete en el uso de instrumentos de medición así como en el redondeo de valores.

(iii) Error de ejecución. Es cualquier discrepancia entre lo que se intentaba hacer y lo que realmente se hace. Por ejemplo, tener en la muestra elementos de una población distinta a la población de interés. Con ciertas precauciones se puede reducir la variabilidad.

1.4 Diferentes intereses y consideraciones respecto de los 'outliers'

Un 'outlier' es una observación que no se ajusta a lo que en un principio constituye para nosotros un conjunto razonable de datos. Surgen dudas subjetivas acerca de la veracidad de esos valores extremos en relación al conjunto de datos obtenido o acerca de la veracidad del modelo de probabilidad inicialmente propuesto, que describe la generación de los datos.

Notemos la importancia del supuesto de un modelo inicial, pues la actitud hacia el conjunto de datos puede diferir ampliamente con modelos diferentes. Por ejemplo, si suponemos que los datos provienen de una distribución Normal puede haber observaciones extremas que nos inquieten, las cuales no llamarían nuestra atención en un modelo con colas más pesadas, digamos una Log-Normal o una Cauchy.

El propósito de tener métodos estadísticos para examinar 'outliers' es, a grandes rasgos, dar una forma de evaluar si nuestra declaración subjetiva de la presencia de 'outliers' en un particular conjunto de datos tendrá implicaciones objetivas importantes más adelante en el análisis de los datos.

La naturaleza de un 'outlier' puede ser aleatoria o determinísta. Es aleatoria cuando no sabemos las causas que lo provocan, es decir, que es inexplicable su presencia. En cambio, es determinísta cuando sabemos por qué se encuentra en el conjunto de datos.

La naturaleza de un 'outlier' puede ser determinísta si su variabilidad es causada por un error de medición o un error de ejecución conocido. Cuando esto sucede el remedio es claro: dichos valores deberán ser removidos de la muestra o reemplazados (volviendo a tomar la observación o corrigiéndola).

Ejemplos (Barnett y Lewis, 1978, 6-10):

1.- En un estudio de probabilidad de temperaturas inferiores durante todos los meses de invierno en las islas británicas, se analizaron gran cantidad de datos de temperaturas tomadas cada hora por varios años y en varios sitios. Las temperaturas fueron registradas en grados Farenheit. Entre los datos tomados en Wick, en el norte de Escocia, se encontraron las siguientes temperaturas correspondientes de la tarde del 31 de diciembre de 1960 a la mañana siguiente del 1 de enero de 1961:

$$43, 43, 41, 41, 41, 42, 43, 58, 58, 41, 41 \tag{1}$$

El primer 58 que corresponde a la media noche y el segundo a la 1.00 a.m. son valores sospechosos, pues están en contraste con las otras observaciones. Sin embargo, se encontró que eran razonables, pues esto se debía a que a media noche el encargado del meteorológico cambió los registros de grados Farenheit a 1/10 de grados Centígrados, por lo que una vez realizada la corrección los valores en grados Farenheit aparecen como sigue:

$$43, 43, 41, 41, 41, 42, 43, 42, 42, 39, 39 \tag{2}$$

Ahora, nótese que aunque no se sospechó de los dos últimos registros en el conjunto (1) eran valores erróneos, pues en vez de ser 41 su valor éra 39.

2.- Hay situaciones en las que los valores registrados son absolutamentes imposibles. En un ejercicio estudiantil había que reportar el número de veces que aparecía el 6 en 10 lanzamientos de un dado, el experimento se realizó 10 veces. Un estudiante reportó los siguientes resultados:

$$2, 0, 3, 12, 2, 0, 1, 1, 3$$

Claramente el valor 12 no es posible. Además por ser 9 observaciones en vez de 10 podría ser que se haya omitido una coma y la secuencia de números quedaría como sigue:

$$2, 0, 3, 1, 2, 2, 0, 1, 1, 3$$

Los ejemplos anteriores ilustran, cómo en estos casos el procesamiento de 'outliers' no necesita un análisis estadístico.

En otras situaciones, se sospechará de alguna(s) observación(es), pero no se podrá dar una explicación tangible del por qué de su valor. En tal contexto se querrá probar de alguna forma si verdaderamente esas observaciones que a 'ojo' nos parecen aberrantes lo son ciertamente.

Es claro que en un conjunto de datos siempre tendremos valores extremos. Lo importante no es que sean extremos, sino que no deban serlo de acuerdo al modelo adoptado.

Esto nos puede sugerir también, que la distribución asumida puede ser incorrecta y sea mejor adoptar otra respecto de la cual el conjunto de datos (incluyendo los 'outliers') aparezca como una muestra aleatoria sin problemas.

3.- Supongamos que la siguiente muestra aleatoria fué obtenida de alguna variable de interés:

$$1.74, 1.46, -1.28, -0.02, -0.04, 0.02, 3.89, 1.35, -0.10, 1.71$$

Deseamos estimar el 'centro' de la población. Inicialmente se supone que la población puede ser Normal, $N(\theta, 1)$. El valor 3.89 que bajo la distribución Normal supuesta es extremo, nos puede hacer sospechar del supuesto de Normalidad. De hecho, los datos fueron generados como una muestra aleatoria de una distribución Cauchy, con función de densidad

$$f(x) = \frac{1}{\pi}(1 + x^2)^{-1}$$

En este caso la media muestral no es consistente y habríamos hecho un uso muy pobre de nuestros datos.

Como hemos visto la prueba de discordancia juega solamente un papel inicial en el análisis de datos, pues después nos interesa ver qué se puede hacer con los 'outliers'. La acción dependerá de nuestros intereses en la situación práctica. Algunas posibilidades como ya vimos son: *rechazarlo* (o reemplazarlo) y continuar con el resto de los datos (o los datos modificados) manteniendo el modelo inicial; modificar el modelo para *incorporar* los 'outliers' de forma que ya no lo sean; *acomodarlos*, es decir, que entren al análisis pero con un peso menor al resto de las observaciones; concentrar la atención en los 'outliers' como una forma de identificar factores desconocidos de importancia práctica.

4.- Supongamos que las observaciones 1.01 y -1.40 son 'outliers' superior e inferior respectivamente, en el siguiente conjunto de 15 residuales de observaciones del semidiámetro vertical de Venus,

-0.30	-0.24	-1.40	+0.18
-0.44	+0.06	-0.22	+0.39
+1.01	+0.63	-0.05	+0.10
+0.48	-0.13	+0.20	

Se puede escoger rechazar las observaciones y continuar el estudio con el resto del conjunto de datos. Pero claro, no podemos estar seguros de que esta acción sea la más adecuada. Tal vez un modelo no-Normal, más sofisticado nos permita *incorporar* las observaciones en una faceta no-aberrante.

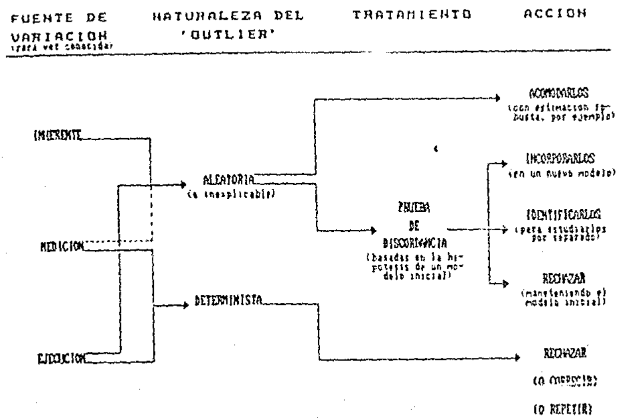
Si se opta por rechazar los 'outliers' es recomendable reportarlos siempre en el estudio, Kruskal (1960).

El rechazo de 'outliers' cuando su naturaleza no es determinística tiene consecuencias estadísticas en el estudio futuro de los datos, pues se trabajará con una muestra reducida. No será una muestra aleatoria, sino una muestra censurada. El hecho de reemplazar los valores rechazados por algún equivalente estadístico (por ejemplo, simulando las observaciones con la distribución asumida para la muestra) tiene consecuencias similiares.

Pero no necesariamente tenemos que empezar el estudio con una prueba de discordancia. Se pueden usar métodos robustos, en los cuáles no se supone ninguna distribución inicial y por lo tanto no es posible una prueba de discordancia.

Desde el punto de vista de los métodos robustos, se planean procedimientos estadísticos los cuáles no examinan directamente los 'outliers', sino que busca *acomodarlos* dándoles una menor influencia en el estudio futuro, lo cual es menos drástico que rechazar completamente los valores, será un rechazo parcial.

La siguiente figura presenta, de una forma simplificada, un diagrama resúmen de la sección.



CAPITULO 2

METODOS BAYESIANOS

2.1 Introducción

Los métodos estadísticos llamados Bayesianos deben su nombre al uso intenso que hacen del resultado –de la teoría matemática de la probabilidad– conocido como teorema de Bayes.

En este enfoque se acepta de manera natural la interpretación de la probabilidad personal [1].

Se trata a los parámetros involucrados como variables aleatorias, aunque sean fijos y desconocidos. En otras palabras, θ es un valor de una variable aleatoria Θ con distribución de probabilidad $f(\theta)$.

Brumat (1977) describe los procedimientos bayesianos de la siguiente manera : "Las técnicas Bayesianas consisten en seleccionar una "opinión inicial" y una "evidencia", en la forma de una distribución inicial y de una función de verosimilitud respectivamente, y trabajando en ellas (usando teorema de Bayes) dar una "opinión final" en la forma de otra distribución de probabilidad."

En efecto, el uso de una distribución inicial, que representa el conocimiento que se tiene acerca de los parámetros desconocidos antes de disponer de los datos, juega un papel importante en el análisis Bayesiano. Tal distribución puede ser usada para representar un conocimiento inicial o una ignorancia relativa. El teorema que combina

[1] La probabilidad personal corresponde a una medida de la seguridad o creencia que el sujeto posee sobre la ocurrencia de un evento con base en su conocimiento, de este modo, un suceso incierto puede tener cualquier valor de probabilidad (dado cierta evidencia E) entre cero y uno. El valor de la probabilidad estará determinado por el conocimiento correspondiente a la persona en cuestión y la evidencia de que disponga.

La interpretación personal de la probabilidad posee las ventajas de otras interpretaciones y pretende eliminar sus desventajas. Para ahondar en el tema se puede consultar Bernett (1979), de Finetti (1978), DeGroot (1970), Valencia y Suárez (1988).

la distribución inicial y los datos para formar la distribución posterior, es precisamente lo que se conoce como el teorema de Bayes.

3.3 Teorema de Bayes

Supongamos que $x = (x_1, \ldots, x_n)$ es un vector de n observaciones cuya distribución de probabilidad $p(x \mid \theta)$ depende de los valores de los parámetros $\theta = (\theta_1, \ldots, \theta_k)$. Supongamos también que θ tiene una distribución de probabilidad $p(\theta)$. Entonces,

$$p(x \mid \theta)p(\theta) = p(x, \theta) = p(\theta \mid x)p(x)$$

Por lo que la distribución condicional de θ dados los datos x es

$$p(\theta \mid x) = \frac{p(x \mid \theta)p(\theta)}{p(x)} \tag{1}$$

El denominador del lado derecho es simplemente la integral (suma) del numerador sobre todos los posibles valores de θ. Aunque el valor de esta integral (suma) depende de los valores observados $x_1, x_2, \ldots, x_n$, no depende de θ y puede ser tratada como una constante cuando es vista como una función de θ. Así,

$$p(\theta \mid x) = cp(x \mid \theta)p(\theta) \tag{2}$$

En la expresión (1) (o equivalentemente (2)) $p(\theta)$, la cual nos habla de lo que se conoce de θ antes de disponer de los datos, es llamada la distribución inicial de θ. De igual forma, $p(\theta \mid x)$, la cual nos habla de lo que se conoce de θ dado el conocimiento de los datos, es llamada la distribución posterior de θ dado x o distribución posterior de θ. El valor c es simplemente una constante de normalización necesaria para asegurar que la función $p(\theta \mid x)$ integre (sume) uno.

En el Apéndice A se enuncia el teorema de Bayes en forma de eventos, de variables aleatorias discretas y de variables aleatorias continuas.

2.3 Verosimilitud

Dados los datos x, $p(x \mid \theta)$ en (1), puede verse no como una función de x, sino de θ. Desde este punto de vista, siguiendo a Fisher (1922), $p(x \mid \theta)$ es llamada la verosimilitud de θ dado x.

El teorema de Bayes nos dice entonces que la distribución de probabilidad para θ posterior a los datos x es proporcional al producto de la distribución inicial de θ y la verosimilitud de θ dado x. Esto es

$$\text{distribución posterior} \propto \text{Verosimilitud} \times \text{distribución inicial}$$

El símbolo de proporcionalidad $\propto$ se usa para indicar que el lado izquierdo es igual al lado derecho, excepto posiblemente por una constante, cuyo valor depende de los valores observados $x_1 \ldots x_n$, pero no depende de θ.

La verosimilitud juega un papel muy importante en la fórmula de Bayes. Es la función a través de la cual, los datos x modifican el conocimiento inicial de θ.

Otro punto importante es que si la verosimiluitd se puede factorizar en la forma $m_1(s,\theta)m_2(x)$ donde s es una función de la muestra, el factor $m_2(x)$ es absorbido en la constante de normalización en el teorema de Bayes (Demostraciones en el Apéndice A).

2.4 Familias Conjugadas de Distribuciones

Aunque el teorema de Bayes puede usarse combinando cualquier distribución inicial con cualquier verosimilitud, es conveniente, especialmente para que el análisis estadístico se facilite, tomar formas especiales de la distribución inicial de tal manera que se simplifiquen los resultados.

Para un modelo $f_X(x \mid \theta)$ podemos buscar una familia de distribuciones iniciales de tal forma que la distribución posterior pertenezca a la misma familia. Dicha familia es llamada *cerrada bajo el muestreo*. La familia también se conoce como *familia conjugada* por la relación especial que debe existir entre esta familia de distribuciones del parámetro y la familia de distribuciones de las observaciones.

Sin embargo, para que tal forma de simplificación sea útil, la familia conjugada de distribuciones debe ser lo suficientemente rica para que nos permita encontrar dentro de la misma familia, en una gran variedad de situaciones, una distribución que represente adecuadamente el conocimiento inicial acerca de θ.

Ahora, se enunciarán las familias conjugadas para muestras de algunas distribuciones en especial.

Comenzaremos con la distribución Normal, en la cual el valor de la precisión o equivalentemente, el valor de la varianza, es conocido.

Teorema.- Supongamos que $x_1, \ldots, x_n$ es una muestra aleatoria de una distribución Normal con media W desconocida y precisición r $(r > 0)$ conocida. Supongamos también que la distribución inicial de W es una Normal con media μ y precisión τ, tal que $-\infty < \mu < \infty$ y $\tau > 0$. Entonces la distribución posterior de W dada la muestra es una distribución Normal con media μ' y precisión $\tau + nr$, donde

$$\mu' = \frac{\tau\mu + nr\overline{x}}{\tau + nr}$$

Demostración.

Para $-\infty < w < \infty$, la función de verosimilitud cumple

$$f(x_1, \ldots, x_n \mid w) \propto \exp[\frac{-r}{2}\sum_{i=1}^{n}(x_i - w)^2]$$

sin embargo,

$$\sum_{i=1}^{n}(x_i - w)^2 = n(w - \overline{x})^2 + \sum_{i=1}^{n}(x_i - \overline{x})^2$$

y como el último término de esta ecuación no depende de w podemos reescribir la verosimilitud como

(3) $$f(x_1, \ldots, x_n \mid w) \propto \exp[-\frac{nr}{2}(w - \overline{x})^2]$$

La fución de distribución inicial de W satisface

(4) $$f(w) \propto \exp[-\frac{\tau}{2}(w - \mu)^2]$$

y así, la posterior de W dada la muestra será proporcional al producto de las funciones especificadas en las relaciones (3) y (4). Pero como

$$\tau(w - \mu)^2 + nr(w - \overline{x})^2 = (\tau + nr)(w - \mu')^2 + \frac{\tau nr(\overline{x} - \mu)^2}{\tau + nr}$$

y ya que el término final no depende de w, puede ser incluido en la constante de proporcionalidad y obtener así

(5) $$f(w \mid x_1, \ldots, x_n) \propto \exp[-\frac{\tau + nr}{2}(w - \mu')^2]$$

con μ' especificado en el enunciado del teorema. De (5) se sigue que la distribución posterior de W es Normal con media μ' y precisión $\tau + nr$. ∎

En las siguientes distribuciones sólo se enunciarán los teoremas y para las demostraciones ó si se desea ahondar en el tema, se pueden consultar DeGroot(1970), DeGroot(1975) y Cox y Hinkley(1974).

Teorema.- Supongamos que $x_1, \ldots, x_n$ es una muestra aleatoria de una distribución Normal con media m $(-\infty < m < \infty)$ conocida y precisión desconocida W. Supongamos también que la distribución inicial de W es una Gama con parámetros α y β tal que $\alpha > 0$ y $\beta > 0$. Entonces, la distribución posterior de W dada la muestra es una Gama con parámetros $\alpha + (n/2)$ y β', donde

$$\beta' = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - m)^2$$

Teorema.- Supongamos que $x_1, \ldots, x_n$ es una muestra aleatoria de una distribución Exponencial con parámetro W desconocido. Supongamos también, que la distribución inicial de W es una Gama con parámetros α y β tal que $\alpha > 0$ y $\beta > 0$. Entonces, la distribución posterior de W dada la muestra es una Gama con parámetros $\alpha + n$ y $\beta + \sum_{i=1}^{n} x_i$.

Teorema.- Supongamos que $x_1, \ldots, x_n$ es una muestra aleatoria de una distribución Poisson con valor de la media W desconocido. Supongamos también que la distribución inicial de W es una Gama con parámetros α y β tal que $\alpha > 0$ y $\beta > 0$. Entonces, la distribución posterior de W dada la muestra es una Gama con parámetros $\alpha + \sum_{i=1}^{n} x_i$ y $\beta + n$.

2.5 Distribuciones de Referencia

Para la aplicación del teorema de Bayes se requiere de una distribución inicial de θ. Como vimos anteriormente, dicha distribución representa el conocimiento del investigador [2] acerca del parámetro θ, antes de disponer de los datos. Sin embargo, podemos tener situaciones en las que no se cuente con un conocimiento inicial y en consecuencia, necesitamos seleccionar una distribución inicial que nos permita matemáticamente obtener una distribución posterior que se base fuertemente en los datos observados (la evidencia) y que influya por lo tanto, lo menos posible, en la distribución posterior. A tal inicial se le conoce como ***distribución inicial de referencia*** (ver por ejemplom, a Bernardo, 1979; DeGroot, 1970). Así, emplear una inicial de referencia, en el sentido práctico, nos hace estar libres de una dificultad teórica.

[2] El investigador es el interesado en el estudio, el sujeto cognoscente.

También puede haber situaciones en las que se disponga de un conocimiento inicial acerca del valor del parámetro θ y sin embargo no queramos usarlo. Esto puede deberse a que se desee hacer un reporte de los resultados (por pensarse que será mas convincente, por ejemplo) analizando los datos con una inicial de referencia, es decir, que nos refleje sólo lo que los datos "dicen" de θ. De esta manera, se deja que los resultados "hablen por sí mismos".

Surge el problema de cómo elegir la distribución inicial con esas características. Históricamente dicho problema es antigüo y hoy día es todavía materia de discusión. En el trabajo original de Bayes, se sugiere tentativamente, que en las situaciones donde el conocimiento acerca de la naturaleza de la distribución inicial fuera insuficiente, ésta puede ser vista como una distribución uniforme. Lo anterior se conoce como postulado de Bayes. Existen muchas objeciones hacia este postulado (Box y Tiao ,1973).

Otra forma de resolver el problema es tomar la distribución inicial de θ como un miembro de la familia conjugada y hacer tender los parámetros de esta distribución a valores tales que no tengan ninguna influencia en la distribución posterior.

2.6 Factor de Bayes

Para contrastar dos hipótesis es conveniente considerar el cociente de momios de una hipótesis, H_1, y de otra hipótesis, H_2. Aquí discutiremos el concepto de cociente de momios. Un cociente de probabilidades posteriores es igual al producto de un cociente de probabilidades iniciales y un cociente de verosimilitudes. Es decir, el teorema de Bayes puede escribirse como:

Cociente de momios posteriores	=	Cociente de momios iniciales	Cociente de verosimilitudes

Esto es,

$$\frac{f(\theta_1 \mid y)}{f(\theta_2 \mid y)} = \frac{f(\theta_1)}{f(\theta_2)} \frac{f(y \mid \theta_1)}{f(y \mid \theta_2)} \tag{6}$$

De aquí se tiene que el cociente de verosimilitudes ó Factor de Bayes es:

$$Factor = \frac{Cociente\ de\ momios\ posteriores}{Cociente\ de\ momios\ iniciales}$$

$$Factor = \frac{f(\theta_1 \mid y)}{f(\theta_2 \mid y)} \frac{f(\theta_2)}{f(\theta_1)} \tag{7}$$

Si el Factor es mayor a 1, se considera que la muestra apoya más la hipótesis H_1 $(\theta = \theta_1)$ que la hipótesis H_2 $(\theta = \theta_2)$. Si el Factor es menor a 1 se considera lo contrario. En Estadística clásica se puede determinar una región de rechazo en términos del cociente de verosimilitudes. Esta región de rechazo, la cual es una región de valores del cociente de verosimilitudes para los cuales H_1 será rechazada en favor de H_2, es de la forma

$$\text{Cociente de verosimilitudes} \leq c$$

donde c es alguna constante. Si c es conocida y el cociente de las verosimilitudes es conocido, se pueden comparar éstos para aceptar o rechazar H_1. Así, c debe ser especificada. En Estadística clásica se considera la siguiente tabla de posibles errores en pruebas de hipótesis

	H_1 verdad	H_1 falsa
Aceptar H_1	NO HAY ERROR	ERROR TIPO II
Rechazar H_1	ERROR TIPO I	NO HAY ERROR

Usualmente se consideran las hipótesis de tal forma que cometer el error tipo I sea más serio que cometer el error tipo II. De esta manera se especifica que la probabilidad de cometer el error tipo I no sea más grande que α (pequeña). La constante c es entonces determinada como:

$$P(\text{error tipo I}) = P(\text{rechazar } H_1 \mid H_1 \text{ es verdadera}) \leq \alpha$$

ya que la región de rechazo es de la forma:

$$\text{Cociente de verosimilitudes} \leq c$$

c es entonces el valor mas grande tal que:

$$P(\text{Cociente de verosimilitudes} \leq c \mid H_1 \text{ es verdadero}) \leq \alpha$$

La relación entre el Factor de Bayes y el Cociente de verosimilitudes clásico proviene de las relaciones (6) y (7).

Esto significa que numéricamente los resultados serán iguales si el cociente de momios iniciales es igual a 1. Esto podría ser pensado como un estado de información inicial difuso, ya que ninguna hipótesis es favorecida. Si el cociente de momios iniciales es diferente de 1, el Factor de Bayes diferirá del Cociente de Verosimilitudes Clásico.De aquí que la inferencia Bayesiana y Clásica no necesariamente sean numéricamente iguales.

Algunos resultados pudieran ser similares numéricamente, pero en términos de interpretación serán diferentes.

Algunas referencias acerca del tema son: Mood y Graybill (1963) y Winkler (1972).

FAMILIA EXPONENCIAL

En este capítulo se presentan algunos resultados de la Familia Exponencial que serán utilizados en los capítulos siguientes.

3.1 Definición

Sea $\{f(x \mid \theta); \theta \in \Omega\}$ una familia de funciones de densidad, donde Ω es el conjunto de intervalos $\Omega = \{\theta : \gamma < \theta < \eta\}$, γ y η son constantes conocidas y

$$(1) \qquad \begin{aligned} f(x \mid \theta) &= \exp\left[\phi(\theta)t(x) + G(\theta) + H(x)\right] \quad a < x < b, \\ &= 0 \quad \text{en otro lado} \end{aligned}$$

Una función de densidad de la forma (1) se dice que es un miembro de la *Familia Exponencial*.

Si además:

i) a y b no dependen de θ y $\gamma < \theta < \eta$,
ii) $\phi(\theta)$ es una función continua no trivial de θ con $\gamma < \theta < \eta$,
iii) $t'(x) \not\equiv 0$ y $H(x)$ son funciones continuas de x para $a < x < b$,

se dice que $f(x \mid \theta)$ es un *miembro regular de la familia exponencial* de funciones de densidad del tipo continuo.

Para un miembro regular de la familia exponencial de funciones del tipo discreto la condición iii) se enuncia:

iii) $t(x)$ es una función no trivial de x para $a < x < b$.

Cuando θ es un vector de dimensión k, entonces se dice que $f(x \mid \theta)$ pertenece a la familia exponencial k-parametral, o bien, que es un miembro de la familia exponencial con k parámetros si

$$f(x \mid \theta) = \exp\left[\sum_{i=1}^{k} \phi_i(\theta) t_i(x) + G(\theta) + H(x)\right]$$

3.2 Resultados de Interés

3.2.1) Si $x_1, \ldots, x_n$ es una muestra aleatoria de tamaño n, de una distribución de la forma (1), entonces $Y = \sum_{i=1}^{n} t(x_i)$ es una estadística suficiente para θ.

Demostración.
La función de distribución conjunta de $x_1, \ldots, x_n$ es

$$\exp\left[\phi(\theta)\sum_{i=1}^{n} t(x_i) + nG(\theta) + \sum_{i=1}^{n} H(x_i)\right] \quad \text{para } a < x_i < b, \gamma < \theta < \eta, i = 1, \ldots, n$$

0 en otro caso
la cual podemos escribir como el producto de dos funciones no negativas

$$\exp\left[\phi(\theta)\sum_{i=1}^{n} t(x_i) + nG(\theta)\right] \exp\left[\sum_{i=1}^{n} H(x_i)\right]$$

y de acuerdo con el teorema de factorización, $Y = \sum_{i=1}^{n} t(x_i)$ es una estadística suficiente para el parámetro θ. ∎

3.2.2) Si $f(x \mid \theta)$ es un miembro regular de la familia exponencial se tiene que

3.2.2.1 $\mathrm{E}(t(x) \mid \theta) = \frac{-G'(\theta)}{\phi'(\theta)} = M(\theta)$

Demostración.

Por definición de función de densidad de probabilidad tenemos

$$\int_a^b f(x \mid \theta) dx = 1$$

$$\int_a^b \exp\left[\phi(\theta)t(x) + G(\theta) + H(x)\right] dx = 1$$

derivando respecto a θ tenemos

$$\frac{\partial}{\partial\theta}\left[\int_a^b exp\left[\phi(\theta)t(x) + G(\theta) + H(x)\right] dx\right] = 0$$

$$(2) \qquad \Longrightarrow \quad \int_a^b \left(t(x)\phi'(\theta) + G'(\theta)\right)\exp\left[\phi(\theta)t(x) + G(\theta) + H(x)\right] dx = 0$$

$$\Longrightarrow \quad \phi'(\theta)\int_a^b t(x)f(x \mid \theta)dx + G'(\theta)\int_a^b f(x \mid \theta)dx = 0$$

$$\Longrightarrow \quad \phi'(\theta)\mathrm{E}\left(t(x) \mid \theta\right) + G'(\theta) = 0$$

$$\Longrightarrow \quad \mathrm{E}\left(t(x) \mid \theta\right) = \frac{-G'(\theta)}{\phi'(\theta)} \quad \blacksquare$$

3.2.2.2 $\mathrm{Var}\left(t(x) \mid \theta\right) = \frac{M'(\theta)}{\phi'(\theta)}$

Demostración.

Derivando (2) respecto a θ se tiene

$$\int_a^b \left[\left(t(x)\phi'(\theta) + G'(\theta)\right)^2 + \left(t(x)\phi''(\theta) + G''(\theta)\right)\right]\exp\left[\phi(\theta)t(x) + G(\theta) + H(x)\right] = 0$$

$$\Longrightarrow \quad \int_a^b \left[\left[t(x)\phi'(\theta)\right]^2 + 2t(x)\phi'(\theta)G'(\theta) + \left[G'(\theta)\right] + t(x)\phi''(\theta) + G''(\theta)\right] f(x \mid \theta)dx = 0$$

$$\Longrightarrow \quad \left[\phi'(\theta)\right]^2\mathrm{E}\left(t^2(x) \mid \theta\right) + \left[2\phi'(\theta)G'(\theta) + \phi''(\theta)\right]\mathrm{E}\left(t(x) \mid \theta\right) + \left[G'(\theta)\right]^2 + G''(\theta) = 0$$

sustituyendo $\mathrm{E}\left(t(x) \mid \theta\right)$ por $\frac{-G'(\theta)}{\phi'(\theta)}$

$$\Longrightarrow \quad \left[\phi'(\theta)\right]^2\mathrm{E}\left(t^2(x) \mid \theta\right) - \left[G'(\theta)\right]^2 - \frac{\phi''(\theta)G'(\theta)}{\phi'(\theta)} + G''(\theta) = 0$$

multiplicando por $\frac{1}{[\phi'(\theta)]^2}$ y despejando obtenemos:

$$\mathrm{E}\left(t^2(x) \mid \theta\right) - \left[\frac{G'(\theta)}{\phi'(\theta)}\right]^2 = \frac{\phi''(\theta)G'(\theta) + G''(\theta)\phi'(\theta)}{\phi'(\theta)\left[\phi'(\theta)\right]^2}$$

$$\Longrightarrow \quad \mathrm{E}\left(t^2(x) \mid \theta\right) - \mathrm{E}^2\left(t(x) \mid \theta\right) = \frac{\phi''(\theta)G'(\theta) - G''(\theta)\phi'(\theta)}{\left[\phi'(\theta)\right]^2}\cdot\frac{1}{\phi'(\theta)}$$

$$\Longrightarrow \quad \mathrm{Var}(t(x) \mid \theta) = \frac{M'(\theta)}{\phi'(\theta)} \quad ■$$

3.2.3) Para una función de densidad $f(x \mid \theta)$ de la forma (1), se tiene que la familia conjugada para la distribución inicial de θ es

$$\mathrm{P}(\theta) \propto \exp\left[\alpha\phi(\theta) + \beta G(\theta)\right] \tag{3}$$

3.2.4) Si $x \mid \theta$ es un miembro de la familia exponencial y la distribución inicial de θ es de la forma (3), entonces la distribución posterior de θ es unimodal.

Demostración.

$$\begin{aligned}
\mathrm{P}(\theta \mid x) &\propto \exp\left\{\phi(\theta)\left[\alpha + \sum_{i=1}^{n} t(x_i)\right] + (n+\beta)G(\theta)\right\} \\
&\propto \exp\left\{\phi(\theta)\sum_{i=1}^{n} t(x_i) + n\left[\left(1+\frac{\beta}{n}\right)G(\theta) + \frac{\alpha}{n}\phi(\theta)\right]\right\}
\end{aligned}$$

Encontrar la moda posterior, es equivalente a encontrar la ecuación de máxima verosimilitud para n observaciones de la familia exponencial

$$\exp\left\{\phi(\theta)t(x) + \left[\left(1+\frac{\beta}{n}\right)G(\theta) + \frac{\alpha}{n}\phi(\theta)\right] + H^*(x)\right\}$$

donde $H^*(x)$ se escoge tal que la integral de esta función respecto a x sea igual a la unidad. Por tanto, para demostrar que es unimodal tenemos que mostrar que el estimador de máxima verosimilitud para una familia exponencial es único.
Sean $x_1, \ldots, x_n$ variables aleatorias independientes e idénticamente distribuidas con densidad

$$\exp\left\{a(\theta)b(x) + c(\theta) + d(x)\right\}$$

La ecuación de verosimilitud resulta ser

$$a'(\hat{\theta})\sum_{i=1}^{n} b(x_i) + nc'(\hat{\theta}) = 0 \tag{4}$$

La segunda derivada del logaritmo de la verosimilitud valuado en $\hat{\theta}$ (solución de (4)) resulta ser negativa ya que es igual a

$$a''(\hat{\theta})\sum_{i=1}^{n} b(x_i) + nc''(\hat{\theta})$$

$$= \left[na''(\hat\theta)\frac{\sum_{i=1}^n b(x_i)}{n} + nc''(\hat\theta)\right]\left[\frac{a'(\theta)}{a'(\theta)}\right]^3$$

$$= -n\left[a''(\hat\theta)a'(\hat\theta)\left(-\frac{\sum_{i=1}^n b(x_i)}{n}\right) - c''(\hat\theta)a'(\hat\theta)\right]\frac{\left[a'(\hat\theta)\right]^2}{\left[a'(\hat\theta)\right]^3}$$

De (4) tenemos $\frac{\sum_{i=1}^n b(x_i)}{n} = -\frac{c'(\hat\theta)}{a'(\hat\theta)}$ así

$$= -n\left[a'(\hat\theta)\right]^2 \left[\frac{a''(\hat\theta)c'(\hat\theta) - c''(\hat\theta)a'(\hat\theta)}{[a'(\hat\theta)]^2}\right]\frac{1}{a'(\hat\theta)}$$

$$= -n[a'(\hat\theta)]^2 \mathrm{Var}(b(x_i) \mid \hat\theta) \le 0$$

Por lo que, excepto en casos degenerados, la solución de (4) es un máximo local, es decir, el estimador de máxima verosimilitud es único. ∎

3.3 Principales Miembros de la Familia Exponencial con 1 parámetro

3.3.1) Bernoulli

$$f(x \mid \theta) = \theta^x(1-\theta)^{1-x}; \qquad x = 0, 1$$

$$\phi(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$$

$$t(x) = x$$

$$G(\theta) = \log(1-\theta)$$

$$H(x) = 0$$

Familia Conjugada para θ: Beta(α, β)

Distribución posterior de θ: Beta$\left(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i\right)$

3.3.2) Binomial

$$
\begin{aligned}
f(x \mid \theta) &= \binom{r}{x}\theta^{x}(1-\theta)^{r-x}; \qquad x = 0,1 \\
\phi(\theta) &= \log\left(\frac{\theta}{1-\theta}\right) \\
t(x) &= x \\
G(\theta) &= r\log(1-\theta) \\
H(x) &= \log\binom{r}{x}
\end{aligned}
$$

Familia Conjugada para θ: Beta(α, β)
Distribución posterior de θ: Beta$(\alpha + rn, \beta + \sum_{i=1}^{n} x_i)$

3.3.3) Poisson

$$
\begin{aligned}
f(x \mid \theta) &= \frac{\theta^{x}\exp(-\theta)}{x!}; \qquad x = 0,1,2,\ldots \\
\phi(\theta) &= \log(\theta) \\
t(x) &= x \\
G(\theta) &= -\theta \\
H(x) &= \log(1/x!)
\end{aligned}
$$

Familia Conjugada para θ: Gama(α, β)
Distribución posterior de θ: Gama$(\alpha + \sum_{i=1}^{n} x_i, \beta + n)$

3.3.4) Exponencial

$$
\begin{aligned}
f(x \mid \theta) &= \theta\exp(-\theta x); \qquad -\infty < x < \infty \\
\phi(\theta) &= -\theta \\
t(x) &= x \\
G(\theta) &= \log(\theta) \\
H(x) &= 0
\end{aligned}
$$

Familia Conjugada para θ: Gama(α, β)
Distribución posterior de θ: Gama$(\alpha + n, \beta + \sum_{i=1}^{n} x_i)$

3.3.5) Normal$(\theta, 1)$

$$\begin{aligned}
f(x \mid \theta) &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x-\theta)^2\right]; \qquad -\infty < x < \infty \\
\phi(\theta) &= \theta \\
t(x) &= x \\
G(\theta) &= \frac{-\theta^2}{2} \\
H(x) &= -\left(\frac{x^2}{2} + \log\sqrt{2\pi}\right)
\end{aligned}$$

Familia Conjugada para θ: Normal(α, β)
Distribución posterior de θ: Normal$\left(\frac{\alpha\beta + n\bar{x}}{\beta + n}, \beta + n\right)$

3.3.5) Normal$(0, \theta)$

$$\begin{aligned}
f(x \mid \theta) &= \left(\frac{\theta}{2\pi}\right)^{1/2} \exp\left(-\frac{1}{2}\theta x^2\right); \qquad -\infty < x < \infty \\
\phi(\theta) &= -\frac{1}{2}\theta \\
t(x) &= x^2 \\
G(\theta) &= \frac{1}{2}\log\left(\frac{\theta}{2\pi}\right) \\
H(x) &= 0
\end{aligned}$$

Familia Conjugada para θ: Gama(α, β)
Distribución posterior de θ: Gama$\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n} x_i^2\right)$

3.3.6) Gama(n, θ)

$$\begin{aligned}
f(x \mid \theta) &= \frac{\theta^n x^{n-1}}{(n-1)!} \exp(-\theta x); \qquad 0 < x < \infty \\
\phi(\theta) &= -\theta \\
t(x) &= x \\
G(\theta) &= n \log\theta + log(n-1) \\
H(x) &= (n-1)\log x
\end{aligned}$$

Algunas referencias acerca del tema son: Bickel (1965), Cox y Hinkley (1974), DeGroot (1975), Hartigan (1956), Hogg (1965), Koopman-Darmois (1936), Lehman (1959) y Lindgren (1968).

CAPITULO 4

UN METODO BAYESIANO

4.1 Introducción

Como ya comenté en el Capítulo 1, se hará la distinción entre observación aberrante y 'outlier'.

El tratamiento de 'outliers', ya sea desde un punto de vista clásico (Barnett y Lewis, 1978, 20-28) o Bayesiano, consta básicamente de los siguientes pasos:

1.- La aplicación de una prueba de 'discordancia' a la(s) observacion(es) más sospechosa(s), que se basa en la adpoción de un modelo inicial al cual pertenece la muestra.

2.- En caso de resultar afirmativa la prueba se procede a tomar una 'acción' contra los 'outliers', que pueden ser, entre otras:

- Incorporarlos usando un nuevo modelo. Por ejemplo, 'outliers' en la distribución Normal pueden ser incorporados en la muestra en una nueva faceta no aberrante, cambiando la distribución adoptada por una t de Student. (ver Capítulo 1)
- Modelarlos directamente, manteniendo el modelo inicial para el resto de las observaciones. (ver Capítulo 1)

El incorporar 'outliers' modificando la distribución adoptada para la muestra por otra que tenga colas mas pesadas, es una forma general de protección contra los 'outliers'. Sin embargo, en situaciones donde tengamos considerable información inicial acerca del modelo que genera 'buenas' observaciones o que genera 'outliers' puede ser preferible modelarlos directamente. El método Bayesiano que aquí se presentará, propuesto por Pettit (1988) utiliza este último enfoque.

Supongamos una muestra aleatoria $x = (x_1, \ldots, x_n)$ de una densidad unimodal $p(x \mid \theta)$, donde θ es un parámetro de localización desconocido, pero algunas observaciones (aberrantes) pueden ser de una densidad diferente. Discutiremos la situación donde sólo existe a lo más una observación aberrante descrita por una traslación en la localización del modelo, esto es, por $p(x \mid \theta + \delta)$ con $\delta > 0$ en caso de tratarse de un 'outlier' superior y $\delta < 0$ si es 'outlier' inferior.

Definimos los siguientes eventos:

$$
\begin{aligned}
A_1 &\equiv \text{'La muestra contiene una observación aberrante'}\\
A_1(i) &\equiv \text{'La observación } x_i \text{ es aberrante, las otras no'}\\
L_i &\equiv \text{'La observación } x_i \text{ es la más grande en la muestra'.}
\end{aligned}
$$

Para que una observación aberrante sea un 'outlier' necesitamos que $p(\delta)$, la especificación inicial que contiene información acerca de la magnitud de la traslación que produce la observación aberrante, sea tal que, para $i = 1, \ldots, n$,

$$Pr(A_1(i) \cap \bar{L}_i) \approx 0 \tag{1}$$

es decir, la componente $p(\delta)$ deberá especificar que con probabilidad alta una observación aberrante será un 'outlier'.

De este modo tenemos

$$
\begin{aligned}
Pr(A_1 \mid x) &= Pr\left(\cup_{i=1}^{n} A_1(i) \mid x\right)\\
&= \sum_{i=1}^{n} \left(Pr(A_1(i) \mid x)\right)\\
&= \sum_{i=1}^{n} Pr\left[(A_1(i) \cap L_i \mid x) \cup (A_1(i) \cap \bar{L}_i \mid x)\right]\\
&= \sum_{i=1}^{n} Pr\left(A_1(i) \cap L_i \mid x\right) + \sum_{i=1}^{n} Pr\left(A_1(i) \cap \bar{L}_i \mid x\right)\\
&= \sum_{i=1}^{n} Pr\left(A_1(i) \mid L_i, x\right) Pr(L_i \mid x) + \sum_{i=1}^{n} Pr\left(A_1(i) \cap \bar{L}_i \mid x\right)
\end{aligned}
$$

si la ecuación (1) se cumple tenemos que si $x_{i^*} = max(x_i)$ entonces $p(L_i \mid x) = 1$ para $i = i^*$ y $p(L_i \mid x) = 0$ para $i \neq i^*$, Así

$$Pr(A_1 \mid x) \approx Pr(A_1(i^*) \mid x) \tag{2}$$

De manera similiar puede hacerse para un 'outlier' inferior.

Lo anterior nos lleva a considerar exclusivamente situaciones que contengan solamente 'outliers' superiores o sólo 'outliers inferiores, y que como ya vimos sean generados por una inflación en el parámetro de escala.

4.3 'Outliers' y la Medida de la Ordenada Predictiva

En una muestra univariada y unimodal hay un orden natural, el cual se ha usado en el presente método para modelar 'outliers'. Para extenderlo a otras situaciones como muestras multivariadas, modelos lineales o muestras de distribuciones que no sean unimodales, se necesita un 'principio de orden', que nos indique el grado en el cual un subconjunto de datos parece aberrante en comparación con las otras observaciones del conjunto de datos. El principio de orden adoptado por Pettit y Smith (1983, 1985) es la Medida de la Ordenada Predictiva (O.P.). Dada una muestra $x = (x_1, \ldots, x_n)$, supongamos que el modelo generador de observaciones no-aberrantes tiene la forma $p(x \mid \psi)$, donde ψ representa todos los parámetros desconocidos. Denotemos por x_S los elementos de x cuyos subíndices están en $S \subset \{1, \ldots, n\}$ y por $x_{(S)}$ los elementos de x cuyos subíndices están en el complemento de S. Entonces, definimos

$$p(x_S \mid x_{(S)}) = \int p(x_S \mid \psi) p(\psi \mid x_{(S)}) d\psi \tag{3}$$

como la ordenada de la densidad predictiva para x_S dado $x_{(S)}$.

Valores pequeños de la ecuación (3) indican que las observaciones x_S son sospechosas en relación a $x_{(S)}$ y a la especificación inicial para ψ. La medida de la Ordenada Predictiva bien puede ser usada en una sóla observación, en pares, en tercias, etc. para comparar su naturaleza aberrante en relación a las otras observaciones.

En algunas distribuciones las observaciones más sospechosas, con respecto a la medida de la Ordenada Predictiva, deberán estar en las observaciones extremas de la muestra. Así, en un modelo en el cual hay exactamente un 'outlier', la observación más sospechosa será el 'outlier'.

Para otras distribuciones, las observaciones más sospechosas no tienen por qué estar en los extremos de la muestra. Por ejemplo, si la distribución es bimodal, una observación del centro de la muestra puede tener la mínima Ordenada Predictiva

Si bien, no necesariamente sucede que la observación con mínima Ordenada Predictiva esté en uno de los extremos de la muestra, síse cumple, bajo ciertas condiciones, para miembros regulares de la Familia Exponencial, con un miembro de la familia conjugada correspondiente como distribución inicial.

Este resultado se muestra en el Teorema de la siguiente sección.

4.4 Teorema

Aunque el resultado será probado para una familia conjugada, es válido para muchas otras distribuciones iniciales. Esto puede checarse para una distribución inicial en particular que fuera de interés en una aplicación. (Pettit, 1988).

Teorema.- Sea $x_1, \ldots, x_n$ una muestra aleatoria de una distribución con densidad $f(x \mid \theta)$, la cuál es un miembro regular de la Familia Exponencial, es decir,

$$f(x \mid \theta) = \exp[\phi(\theta)t(x) + G(\theta) + H(x)] \quad a < x < b, \quad \gamma < \theta < \eta$$

la familia conjugada para θ:

$$p(\theta) \propto \exp[\alpha\phi(\theta) + \beta G(\theta)]$$

Sea $\phi(.)$ una función monótona y $t(.)$ un incremento monótono y positivo. Entonces

(a) Si $f(x \mid \theta)$ es unimodal con moda en (a, b), entonces $p(x_i \mid x_{(i)})$ toma su valor mínimo en la más pequeña ó en la más grande de las observaciones.

(b) Si $f(x \mid \theta)$ toma su valor máximo en a (b) y es monótona, entonces $p(x_i \mid x_{(i)})$ toma su valor mínimo en la más grande (pequeña) de las observaciones.

Demostración.
Antes de pasar a la demostración probaremos el siguiente Lema:

Lema.- $p(x_i \mid x_{(i)}) < p(x_j \mid x_{(j)})$ si y sólo si $p(x_i \mid \hat{\theta}_{ij}) < p(x_j \mid \hat{\theta}_{ij})$ donde $\hat{\theta}_{ij}$ es un valor único tal que $p(\hat{\theta}_{ij} \mid x_{(i)}) = p(\hat{\theta}_{ij} \mid x_{(j)})$.

Prueba.

$$\begin{aligned} p(x_i \mid x_{(i)}) &= \frac{p(x_i, x_{(i)})}{p(x_{(i)})} \\ &= \frac{p(x)}{p(x_{(i)})} \end{aligned}$$

multiplicando y dividiendo por $p(\theta)\prod_{k=1}^{n} p(x_k \mid \theta)$ tenemos

$$= \frac{p(\theta)\prod_{k=1}^{n} p(x_k \mid \theta)}{p(\theta)\prod_{k=1}^{n} p(x_k \mid \theta)} \frac{p(x)}{p(x_{(i)})}$$

agrupando

$$= \left[\frac{p(\theta)\prod_{k=1}^{n} p(x_k \mid \theta)}{p(x_{(i)})}\right]\left[\frac{p(\theta)\prod_{k=1}^{n} p(x_k \mid \theta)}{p(x)}\right]^{-1}$$

como $\prod_{k=1}^{n} p(x_k \mid \theta) = p(x \mid \theta)$

$$= \left[\frac{p(\theta)p(x_i \mid \theta)\prod_{k\neq i} p(x_k \mid \theta)}{p(x_{(i)})}\right]\left[\frac{p(\theta)p(x \mid \theta)}{p(x)}\right]^{-1}$$

utilizando que $\prod_{k\neq i} p(x_k \mid \theta) = p(x_{(i)} \mid \theta)$ tenemos

$$= \left[p(x_i \mid \theta)\frac{p(\theta)p(x_{(i)} \mid \theta)}{p(x_{(i)})}\right]\left[\frac{p(\theta)p(x \mid \theta)}{p(x)}\right]^{-1}$$

$$= \frac{p(x_i \mid \theta)p(\theta \mid x_{(i)})}{p(\theta \mid x)}$$

Por lo que

$$\frac{p(x_i \mid x_{(i)})}{p(x_j) \mid x_{(j)})} = \frac{p(x_i \mid \theta)p(\theta \mid x_{(i)})}{p(x_j \mid \theta)p(\hat{\theta} \mid x_{(j)})}$$

Así, para probar el Lema sólo debemos mostrar que $\hat{\theta}_{ij}$ es único. Como

$$\begin{aligned}
p(\theta \mid x_{(i)}) &\propto p(\theta)p(x_{(i)} \mid \theta) \\
&\propto \exp\left[\alpha\phi(\theta) + \beta G(\theta)\right]\exp\left[\sum_{k\neq i}\phi(\theta)t(x_k) + (n-1)G(\theta) + \sum_{k\neq i} H(x_k)\right] \\
&\propto \exp\left\{\phi(\theta)\left[\sum_{k\neq i} t(x_k) + \alpha\right] + (n+\beta-1)G(\theta)\right\}\exp\left[\sum_{k\neq i} H(x_k)\right] \\
&= \exp\left\{\phi(\theta)\left[\alpha + \sum_{k\neq i} t(x_k)\right] + (n+\beta-1)G(\theta)\right\} C_i^{-1}
\end{aligned}$$

donde

$$C_i = \int \exp\left\{\phi(\theta)\left[\alpha + \sum_{k\neq i} t(x_k)\right] + (n+\beta-1)G(\theta)\right\} d\theta$$

por lo que

$$p(\hat{\theta} \mid x_{(i)}) = p(\hat{\theta} \mid x(j))$$

$\Longleftrightarrow$

$$\frac{\exp\left\{\phi(\hat{\theta})\left[\alpha+\sum_{k\neq j} t(x_k)\right]+(n+\beta-1)G(\hat{\theta})\right\}}{\exp\left\{\phi(\hat{\theta})\left[\alpha+\sum_{k\neq i} t(x_k)\right]+(n+\beta-1)G(\hat{\theta})\right\}}=\frac{C_j}{C_i}$$

$\Longleftrightarrow$

$$\frac{\exp\left[\phi(\hat{\theta})\left[\alpha+\sum_{k\neq j} t(x_k)\right]\right]}{\exp\left[\phi(\hat{\theta})\left[\alpha+\sum_{k\neq i} t(x_k)\right]\right]}=\frac{C_j}{C_i}$$

$\Longleftrightarrow$

$$\exp\left\{\phi(\hat{\theta})\left[\alpha+\sum_{k\neq j} t(x_k)-\alpha-\sum_{k\neq i} t(x_k)\right]\right\}=\frac{C_j}{C_i}$$

$\Longleftrightarrow$

$$\phi(\hat{\theta})\left[t(x_i)-t(x_j)\right]=\log(C_j/C_i)$$

$\Longleftrightarrow$

$$\phi(\hat{\theta})=\left[t(x_i)-t(x_j)\right]^{-1}\log(C_j/C_i)$$

Como $\phi(.)$ es monótona creciente hay una única solución a esta ecuación y por lo tanto $\hat{\theta}_{ij}$ es único. $\circ$

Usando este Lema podemos ver que si etiquetamos las observaciones, por conveniencia, de tal manera que $x_1 < \ldots < x_n$, entonces podemos saber si $p(x_1 \mid x_{(1)}) < p(x_n \mid x_{(n)})$ ó si $p(x_1 \mid x_{(1)}) > p(x_n \mid x_{(n)})$ comparando $p(x_1 \mid \hat{\theta}_{1n})$ y $p(x_n \mid \hat{\theta}_{1n})$. De este modo, para probar el Teorema sólo debemos mostrar que $p(x_i \mid x_{(i)})$, para $1 \leq i \leq n$, no puede ser menos que $p(x_1 \mid x_{(1)})$ y al mismo tiempo ser menor que $p(x_n \mid x_{(n)})$.

Para el inciso (b) del Teorema es inmediato, ya que $f(x \mid \theta)$ toma su valor máximo en a (b) y por ser monótona $p(x_n \mid \theta) < p(x_i \mid \theta)$ $(p(x_1) \mid \theta) < p(x_i \mid \theta))$ para todo θ, en particular para $\theta = \hat{\theta}_{in}$ $(\theta = \hat{\theta}_{1i})$ y usando el Lema tenemos que $p(x_n \mid x_{(n)}) < p(x_i \mid x_{(i)})$ $(p(x_1 \mid x_{(1)}) < p(x_i \mid x_{(i)}))$.

Para el inciso (a) usaremos el resultado 3.2.4 del Capítulo 3, que nos dice: "Si $x \mid \theta$ es un miembro de la Familia Exponencial y la distribución inicial de θ pertenece a la familia conjugada, entonces la distribución posterior de θ es unimodal ".

Con esto tenemos que si $p(\theta \mid x_{(i)})$ tiene moda en $\tilde{\theta}_i$ entonces $\hat{\theta}_{ij}$ debe estar entre $\tilde{\theta}_i$ y $\tilde{\theta}_j$ (Pettit, 1988).

Finalmente para probar el Teorema debemos mostrar que $\hat{\theta}_{in} < \hat{\theta}_{1i}$ para que entonces, por el hecho de que $f(x \mid \theta)$ es unimodal tengamos que

$$p(x_i \mid \hat{\theta}_{1i}) < p(x_1 \mid \hat{\theta}_{1i}) \Longrightarrow p(x_n \mid \hat{\theta}_{in}) < p(x_i \mid \hat{\theta}_{in})$$

Ahora, ya que $\hat{\theta}_{ij}$ debe estar entre $\tilde{\theta}_i$ y $\tilde{\theta}_j$ es suficiente mostrar que $\tilde{\theta}_n < \tilde{\theta}_i < \tilde{\theta}_1$, para que así $\hat{\theta}_{in} < \hat{\theta}_{1i}$.

Como $\tilde{\theta}_i$ es moda de $p(\theta \mid x_{(i)})$, es solución de (ver resultado **3.2.4** del Capítulo 3)

$$\phi'(\theta)\left[\alpha + \sum_{k\neq i} t(x_k)\right] + (n+\beta-1)G'(\theta) = 0$$

$\Longrightarrow$

$$\frac{\phi'(\theta)}{G'(\theta)} = \frac{-(n+\beta-1)}{\alpha + \sum_{k\neq i} t(x_k)}$$

por lo cual

$$\frac{\phi'(\tilde{\theta}_1)}{G'(\tilde{\theta}_1)} = \frac{-(n+\beta-1)}{\alpha + \sum_{k\neq 1} t(x_k)}$$
$$\frac{\phi'(\tilde{\theta}_i)}{G'(\tilde{\theta}_i)} = \frac{-(n+\beta-1)}{\alpha + \sum_{k\neq i} t(x_k)}$$
$$\frac{\phi'(\tilde{\theta}_n)}{G'(\tilde{\theta}_n)} = \frac{-(n+\beta-1)}{\alpha + \sum_{k\neq n} t(x_k)}$$

y como $t(x)$ es un incremento monótono se tiene

$$\frac{\phi'(\tilde{\theta}_1)}{G'(\theta_1)} > \frac{\phi'(\tilde{\theta}_i)}{G'(\theta_i)} > \frac{\phi'(\tilde{\theta}_n)}{G'(\theta_n)}$$

y por el resultado **3.3.2** del Capítulo 3, tenemos

$$\mathrm{E}(t(x) \mid \tilde{\theta}_n) < \mathrm{E}(t(x) \mid \tilde{\theta}_i) < \mathrm{E}(t(x) \mid \tilde{\theta}_1)$$

lo que nos implica que

$$\tilde{\theta}_n < \tilde{\theta}_i < \tilde{\theta}_1$$

por ser t(.) positivo. Quedando así demostrado el Teorema. ∎

Por el uso en la demostración de un orden entre las $\tilde{\theta}_i$ y las $\hat{\theta}_{ij}$ tenemos que la distribución $f(x \mid \theta)$ debe pertenecer a la Familia Exponencial con 1 parámetro, es decir, θ no es un vector de parámetros.

4.5 Observaciones

A lo largo del capítulo hemos mencionado ciertas condiciones que deben darse para que el método propuesto por Pettit (1988) pueda ser utilizado. A continuación haremos un breve resúmen de esas condiciones:

1.- Las observaciones aberrantes deben ser 'outliers', lo que equivale a pedir que las observaciones aberrantes sean observaciones extremas [1]. En el caso de tratarse de un 'outlier' superior, es equivalente pedir que la ecuación (2) se cumpla. Esta condición en la realidad puede ser muy fuerte, pudiendo ser más razonable el pedir que esté entre las 2 ó 3 más grandes.

2.- Debemos tener suficiente información inicial acerca del modelo que genera 'buenas' observaciones y del que genera 'outliers'. El modelo generador de 'outliers' deber ser igual al modelo que genera 'buenas' observaciones, pero con una traslación en su localización.

3.- Sólo podemos considerar situaciones en las que tengamos sólo 'outliers' superiores ó situaciones que contengan sólo 'outliers' inferiores.

4.- El modelo considerado $f(x \mid \theta)$, debe ser un miembro de la Familia Exponencial con 1-parámetro, donde θ es un parámetro de localización.

Si se cumplen las condiciones anteriores, o si las aproximaciones hechas son razonables, usar el método es un camino muy útil ya que es una forma de detectar observaciones sospechosas en muchos contextos (Pettit, 1988).

Además, si nuestro conocimiento acerca del fenómeno nos indica que las condiciones anteriores se cumplen, el método es una buena forma de aprovechar dicho conocimiento, en vez de desperdiciarlo.

La idea general de que una observación aberrante es aquella generada por un mecanismo diferente del que genera la mayoría de las observaciones, así sola, sin más elaboración, no engloba la noción de "distanciamiento", que es lo que se entiende por el término 'outlier'. Por la forma del método aquí presentado, esta dificultad desaparece.

Además creemos que el método es muy versátil. Aquí se ha hecho su presentación bajo ciertas consideraciones para facilitar su exposición, pero se puede extender a muchos otros casos. Por ejemplo, se pueden contemplar situaciones en las que se tengan tanto 'outliers' inferiores como superiores.

En el capítulo siguiente mediante los ejemplos desarrollados se verá en una forma más clara la utilidad de su aplicación.

Por último una observación acerca de la medida de la Ordenada Predictiva que hace Geisser (1980) es que no es invariante bajo transformaciones.

[1] Las más grandes si se trata de 'outliers' superiores y las más pequeñas si es de 'outliers' inferiores.

EJEMPLOS

5.1 Distribución Exponencial

5.1.1) Modelo Exponencial

Asumimos que en una muestra de tamaño n, la mayoría de las observaciones tienen una distribución Exponencial con función de densidad $f(x \mid \lambda) = \lambda \exp(-\lambda x)$, $x > 0$. Consideremos la distribución inicial de λ como una Gama con parámetros α y β (por ser la Familia Conjugada). Utilizando el Teorema de la sección **4.4** del Capítulo 4, vemos que la mínima Ordenada Predictiva para la distribución debe estar en la observación más grande. De este modo, se asumirá que sólo unas pocas observaciones aberrantes pueden tener una distribución Exponencial con función de densidad

$$g_s(x \mid \lambda) = \delta\lambda \exp(-\delta\lambda x) \quad x > 0,\ 0 < \delta < 1$$

Notemos que bajo este modelo sólo son detectados 'outliers' superiores. Esto es debido a que en la distribución exponencial las observaciones grandes son menos "esperadas" que las observaciones chicas, en el sentido de que si $x_1 < x_2$ entonces $Pr(x_1 - \epsilon < X < x_1 + \epsilon) > Pr(x_2 - \epsilon < x_2 + \epsilon)$, y por esto es razonable que max x_i tenga la mínima Ordenada Predictiva.

Este es el modelo general que desarrollaremos, primero considerando δ conocida y después discutiremos asignándole una distribución inicial.

5.1.2) δ Conocida

Para la aplicación del modelo tenemos que checar (ver Capítulo 4) la aproximación del inciso (2) sección **4.2**. Es decir, debemos tener que con probabilidad alta una observación aberrante será la más grande en la muestra. Esta probabilidad es

$$\frac{\Gamma(n)\Gamma(\delta+1)}{\Gamma(n+\delta)} \tag{1}$$

(Ver Apéndice B). Esta probabilidad es igual a $1/n$ para $\delta = 1$ y tiende a la unidad cuando $\delta \to 0$.

Por conveniencia se asumirá que los valores de la muestra son $z_1, \ldots, z_n$ donde z_n es la observación más grande.

Denotando por

$H_0 \equiv$ 'No hay 'outliers' en la muestra'

tenemos

$$f(\lambda) \propto \lambda^{\alpha-1} \exp(-\beta\lambda)$$

$$f(\underline{z} \mid H_0, \lambda) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} z_i)$$

con lo cual

$$\begin{aligned} f(\lambda \mid \underline{z}, H_0) &\propto f(\lambda) f(\underline{z} \mid H_0, \lambda) \\ &\propto \lambda^{n+\alpha-1} \exp\left[-\lambda\left(\sum_{i=1}^{n} z_i + \beta\right)\right] \end{aligned}$$

introduciendo la constante de proporcionalidad se tiene

$$f(\lambda \mid \underline{z}, H_0) = \frac{(\sum_{i=1}^{n} z_i + \beta)^{n+\alpha}}{\Gamma(n+\alpha)} \lambda^{n+\alpha-1} \exp\left[-\lambda\left(\sum_{i=1}^{n} z_i + \beta\right)\right]$$

Así, la distribución posterior de λ dado que en la muestra no hay 'outliers' es una Gama con parámetros $n+\alpha$ y $\sum_{i=1}^{n} z_i + \beta$.

De igual forma, denotando

$H_1 \equiv$ 'Hay un 'outlier' en la muestra'

tenemos

$$f(\lambda) \propto \lambda^{\alpha-1} \exp(-\beta\lambda)$$

$$f(\underline{z} \mid H_1, \lambda) = \lambda^n \delta \exp\left[-\lambda\left(\sum_{i=1}^{n-1} z_i + \delta z_n\right)\right]$$

Así,

$$f(\lambda \mid \underline{z}, H_1) \propto \lambda^{n+\alpha-1} \delta \exp\left[-\lambda\left(\sum_{i=1}^{n-1} z_i + \delta z_n + \beta\right)\right]$$

introduciendo la constante de proporcionalidad

$$f(\lambda \mid \underline{z}, H_1) = \frac{\left(\sum_{i=1}^{n-1} z_i + \delta z_n + \beta\right)^{n+\alpha}}{\Gamma(n+\alpha)} \lambda^{n+\alpha-1} \exp\left[-\lambda\left(\sum_{i=1}^{n-1} z_i + \delta z_n + \beta\right)\right]$$

Por lo que la distribución posterior de λ dado que en la muestra hay un 'outlier', es un Gama con parámetros $n+\alpha$ y $\sum_{i=1}^{n-1} z_i + \delta z_n + \beta$.

La probabilidad de que la muestra no contenga 'outliers' es

$$\begin{aligned} f(H_0, \underline{z}) &= f(\underline{z} \mid H_0) f(H_0) \\ &= \left[\frac{f(\lambda) f(\underline{z} \mid H_0, \lambda)}{f(\lambda \mid H_0, \underline{z})} \right] f(H_0) \\ &= \left[\Gamma(n+\alpha) \left(\sum_{i=1}^{n} z_i + \beta \right)^{-(n+\alpha)} \right] f(H_0) \end{aligned}$$

La probabilidad de que la muestra contenga un 'outlier' es

$$\begin{aligned} f(H_1, \underline{z}) &= f(\underline{z} \mid H_1) f(H_1) \\ &= \left[\frac{f(\lambda) f(\underline{z} \mid H_1, \lambda)}{f(\lambda \mid H_1, \underline{z})} \right] f(H_1) \\ &= \left[\Gamma(n+\alpha)\delta \left(\sum_{i=1}^{n-1} z_i + \delta z_n + \beta \right)^{-(n+\alpha)} \right] f(H_1) \end{aligned}$$

Usando el Factor de Bayes para comparar el modelo en el cual la muestra no contiene 'outliers' contra el de que contiene un 'outlier' es

$$\begin{aligned} Factor &= \frac{f(H_0 \mid \underline{z})}{f(H_1 \mid \underline{z})} \frac{f(H_1)}{f(H_0)} \\ &= \frac{f(\underline{z} \mid H_0)}{f(\underline{z} \mid H_1)} \\ &= \frac{\Gamma(n+\alpha) \left(\sum_{i=1}^{n} z_i + \beta\right)^{-(n+\alpha)}}{\Gamma(n+\alpha)\delta \left(\sum_{i=1}^{n-1} z_i + \delta z_n + \beta\right)^{-(n+\alpha)}} \qquad (2) \\ &= \delta^{-1} \left(\sum_{i=1}^{n} z_i + \beta \right)^{-(n+\alpha)} \left(\sum_{i=1}^{n-1} z_i + \delta z_n + \beta \right)^{(n+\alpha)} \\ &= \delta^{-1} \left[1 + \frac{z_n(\delta - 1)}{\sum_{i=1}^{n} z_i + \beta} \right]^{n+\alpha} \end{aligned}$$

Siendo entonces el Factor de Bayes una función de $z_n \left(\sum_{i=1}^{n} z_i + \beta\right)^{-1}$. Si hacemos $\alpha, \beta \to 0$, representando un vago conocimiento inicial, el Factor de Bayes será una función de la estadística de prueba clásica del cociente de verosimilitudes, $z_n \left(\sum_{i=1}^{n} z_i\right)^{-1}$, propuesta por Fisher en 1929 (Barnett, 1978). A continuación veremos algunos ejemplos numéricos.

Ejemplo 1

Supongamos que la distribución exponencial de la que proviene la mayoría de las observaciones de la muestra es $f(x \mid \lambda) = \lambda \exp(-\lambda x)$. Elegimos el valor de λ igual a 10. Generamos 99 observaciones de dicha distribución usando el paquete *STATGRAPHICS*. Suponiendo una $\delta = .1$ para el 'outlier', generamos de igual forma a través de *STATGRAPHICS*, una observación, siendo esta igual a 0.459.

La muestra de tamaño 100 ya ordenada es

0.001	0.001	0.001	0.003	0.005
0.006	0.007	0.007	0.009	0.009
0.010	0.010	0.011	0.012	0.012
0.014	0.017	0.017	0.018	0.019
0.024	0.026	0.027	0.027	0.027
0.028	0.028	0.029	0.031	0.031
0.032	0.032	0.035	0.037	0.038
0.039	0.039	0.040	0.042	0.043
0.046	0.048	0.050	0.053	0.053
0.063	0.064	0.064	0.064	0.066
0.066	0.067	0.068	0.069	0.071
0.072	0.074	0.077	0.078	0.083
0.087	0.088	0.088	0.089	0.096
0.096	0.096	0.101	0.102	0.105
0.111	0.114	0.118	0.128	0.129
0.130	0.131	0.132	0.134	0.140
0.144	0.146	0.152	0.179	0.203
0.208	0.239	0.250	0.261	0.262
0.271	0.280	0.293	0.299	0.305
0.329	0.334	0.368	0.398	0.459

Como elegimos $\lambda = 10$, suponemos la distribución inicial de λ como una Gama con $\alpha = 5$ y $\beta = 0.5$, pues la media de la distribución será precisamente el valor de λ, con una varianza de 20.

El Factor de Bayes calculado mediante la ecuación (2) resulta ser:

$$Factor = 0.132$$

Sabemos que el valor de este Factor es siempre mayor que cero; apoyando la hipótesis H_0 si dicho valor es menor que uno. Siendo este apoyo más claro conforme tienda a cero. Es indiferente entre las dos hipótesis si el valor es 1 y apoya más la hipótesis H_1 si es mayor que uno (ver sección **2.6** del Capítulo 2).

En este caso, como el valor del Factor de Bayes es menor que 1, nos apoya la hipótesis de que el valor 0.459 es un 'outlier' .

Ejemplo 2

Utilizando los mismos parámetros, pero ahora suponiendo $\delta = .01$ para el 'outlier'la observación generada a través del *STATGRAPHICS* resulta ser 3.978 y el Factor de Bayes para este caso es:

$$Factor = 3.6 \times 10^{-14}$$

Por tanto, como el Factor es cercano a cero nos apoya que el valor 3.978 es un 'outlier'.

Vemos que en el segundo ejemplo es "más claro" el rechazo de la hipótesis de que 'no hay outliers', esto se debe en parte a que la probabilidad de que una observación aberrante sea el más grande en la muestra es mayor con $\delta = .01$. La siguiente tabla dada por Pettit(1988) muestra algunos valores de esta probabilidad (expresión (1)) para distintos valores de n y δ.

Tabla 1

Valores de la ecuación (1) para distintos valores de n y δ.

n	δ				
	0.1	0.08	0.05	0.03	0.01
5	0.817	0.850	0.902	0.940	0.979
10	0.759	0.801	0.870	0.919	0.972
15	0.728	0.775	0.852	0.908	0.968
20	0.707	0.757	0.839	0.900	0.965
30	0.678	0.732	0.822	0.889	0.961
40	0.659	0.715	0.810	0.881	0.958
50	0.644	0.702	0.801	0.875	0.956
75	0.618	0.680	0.785	0.864	0.952
100	0.600	0.664	0.773	0.857	0.950

Consideremos ahora la posibilidad de tener dos 'outliers' superiores, x_{n-1} y x_n. Procediendo de igual forma, denotamos

$H_2 \equiv$ 'Hay dos 'outliers' en la muestra'

Así

$$f(\lambda) \propto \lambda^{\alpha-1} \exp(-\beta\lambda)$$

$$f(\underline{x} \mid H_2, \lambda) = \lambda^n \delta^2 \exp\left[-\lambda\left(\sum_{i=1}^{n-2} x_i + \delta(x_n + x_{n-1})\right)\right]$$

Por lo que,

$$f(\lambda \mid \underline{z}, H_2) \propto \lambda^{n+\alpha-1}\delta^2 \exp\left[-\lambda\left(\sum_{i=1}^{n-2} z_i + \delta(z_n + z_{n-1}) + \beta\right)\right]$$

introduciendo la constante de proporcionalidad

$$= \frac{\left(\sum_{i=1}^{n-2} z_i + \delta(z_n + z_{n-1}) + \beta\right)^{n+\alpha}}{\Gamma(n+\alpha)}\lambda^{n+\alpha-1} \exp\left[-\lambda\left(\sum_{i=1}^{n-2} z_i + \delta(z_n + z_{n-1}) + \beta\right)\right]$$

Siendo entonces una Gama con parámetros $n+\alpha$ y $\sum_{i=1}^{n-2} z_i + \delta(z_n + z_{n-1}) + \beta$. Lo que nos implica

$$\begin{aligned} f(H_2, \underline{z}) &= f(\underline{z} \mid H_2) f(H_2) \\ &= \left[\Gamma(n+\alpha)\delta^2\left(\sum_{i=1}^{n-2} z_i + \delta(z_n + z_{n-1}) + \beta\right)^{-(n+\alpha)}\right] f(H_2) \end{aligned}$$

Así, el Factor de Bayes, que nos compara el modelo de no 'outliers' con el de dos 'outliers' es

$$\begin{aligned} Factor &= \delta^{-2}\left(\sum_{i=1}^{n} z_i + \beta\right)^{-(n+\alpha)}\left(\sum_{i=1}^{n-2} z_i + \delta(z_n + z_{n-1}) + \beta\right)^{(n+\alpha)} \\ &= \delta^{-2}\left[1 + \frac{(z_n + z_{n-1})}{\sum_{i=1}^{n} z_i + \beta}(\delta - 1)\right]^{n+\alpha} \end{aligned} \tag{3}$$

Y es de este modo una función de $(z_n + z_{n-1})(\sum_{i=1}^{n} z_i + \beta)^{-1}$.

En este caso la prueba estadística clásica correspondiente al cociente de verosimilitudes es $(z_n + z_{n-1})(\sum_{i=1}^{n} z_i)^{-1}$.

Barnett y Lewis (1978) y Kimber (1982) hacen notar que esta estadística es susceptible a efectos engañosos. Si z_n es muy grande, entonces el par z_n y z_{n-1} pueden ser declarados como 'outliers' aunque z_{n-1} parezca ser un valor razonable.

Considerando la relación existente entre el Factor de Bayes y el cociente de verosimilitudes, ¿habrá un fenómeno similar que afecte la inferencia Bayesiana?

Sí existe, pero lo podemos resolver de la siguiente manera:

El Factor de Bayes del modelo de un 'outlier' contra el de dos 'outliers' es

$$\begin{aligned} Factor &= \delta^{-1}\left(\sum_{i=1}^{n-1} z_i + \delta z_n + \beta\right)^{-(n+\alpha)}\left(\sum_{i=1}^{n-2} z_i + \delta(z_n + z_{n-1}) + \beta\right)^{(n+\alpha)} \\ &= \delta^{-1}\left[1 + \frac{z_{n-1}}{\sum_{i=1}^{n-1} z_i + \delta z_n + \beta}(\delta - 1)\right]^{n+\alpha} \end{aligned} \tag{4}$$

Siendo una función de $x_{n-1}(x_1 + \ldots + x_{n-1} + \delta x_n + \beta)^{-1}$.

Esto nos asegura que, si bien un valor grande de x_n puede llevarnos a tener poca probabilidad para no 'outlier' contra dos 'outliers', es necesario un valor grande de x_{n-1} para tener gran peso para dos 'outliers' contra un 'outlier'.

De igual forma podemos tener que si bien un valor grqnde de x_{n-1} puede causarnos gran peso para no 'outliers' contra un 'outlier', tenemos poco peso para no 'outliers' contra dos 'outliers'.

Ejemplo 3

En la muestra de los ejemplos anteriores supongamos que el último valor (x_{100}) es 8.055, el cual fué generado con una $\delta = .01$ por medio del paquete *STATGRAPHICS*. En este caso el valor del Factor de Bayes que nos compara el modelo de no 'outliers' contra el de dos 'outliers'(ecuación (3)) es

$$Factor = 7.4 \times 10^{-25}$$

que por ser cercano a cero apoya la hipótesis de que hay dos 'outliers' en la muestra, aunque la observación x_{99} sea un valor razonable.

Utilizando el Factor de Bayes para comparar el modelo de un'outlier' contra el de dos 'outlier' (ecuación (4)), tenemos

$$Factor = 1.403$$

Este valor por ser mayor que 1 nos apoya más la hipótesis de que sólo hay un 'outlier', ya que el valor 0.398 es un valor razonable.

5.1.3) δ Desconocida

Si δ es desconocida debemos especificar una distribución conjunta inicial para λ y δ. Si se tienen $n-1$ observaciones $x_1, \ldots, x_{n-1}$ de la distribución Exponencial $f(x \mid \lambda) = \lambda \exp(-\lambda x)$, $x > 0$ y una observación, x_n, de la distribución Exponencial $g(x \mid \lambda) = \delta\lambda \exp(-\delta\lambda x)$, entonces la verosimilitud es:

$$\lambda^{n-1}\lambda\delta \exp\left[-\lambda \sum_{i=1}^{n-1} x_i - \lambda\delta x_n\right]$$

Que en la forma general de la Familia Exponencial queda así,

$$\exp\left[-\lambda \sum_{i=1}^{n-1} x_i - \lambda\delta x_n + (n-1)\log\lambda + \log(\lambda\delta)\right]$$

lo cual nos sugiere utilizar una conjugada inicial como

$$p(\lambda, \delta) \propto exp\left[-b\lambda - e\lambda\delta + a\log\lambda + f\log(\lambda\delta)\right]$$

donde la constante de proporcionalidad está dada por el inverso multiplicativo de:

$$\int_0^\infty \int_0^\infty \exp[-b\lambda - e\lambda\delta + a\log\lambda + f\log(\lambda\delta)]d\delta d\lambda$$
$$= \int_0^\infty \lambda^f \lambda^a \exp(-b\lambda)\left[\int_0^\infty \delta^f \exp(-e\lambda\delta)d\delta\right]d\lambda$$

Utilizando que $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\exp(-x)dx$ y que $\int_0^\infty x^{\alpha-1}\exp(-\lambda x)dx = \Gamma(\alpha)/\lambda^\alpha$ (ver Hoel 19 ,129) tenemos

$$= \int_0^\infty \lambda^f \lambda^a \exp(-b\lambda)\frac{\Gamma(f+1)}{(e\lambda)^{f+1}}d\lambda$$
$$= \frac{\Gamma(f+1)\Gamma(a)}{e^{f+1}b^a}$$

Y de esta forma la constante de proporcionalidad es

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

La distribución marginal inicial para λ es

$$\int_0^\infty \frac{b^a e^{f+1}}{\Gamma(f+1)\Gamma(a)}\lambda^a\lambda^f\delta^f \exp(-b\lambda - e\lambda\delta)d\delta$$

$$= \frac{b^a e^{f+1}}{\Gamma(f+1)\Gamma(a)}\lambda^a\lambda^f \exp(-b\lambda)\int_0^\infty \delta^f\exp(-e\lambda\delta)d\delta$$
$$= \frac{b^a}{\Gamma(a)}\lambda^{a-1}\exp(-b\lambda)$$

Esta es una Gama con parámetros a y b.
La distribución marginal inicial para δ es Siegel (ver Apéndice C) de la forma

$$p(\delta) = \Gamma(a+f+1)\delta^f\left\{\Gamma(a)\Gamma(f+1)(b/e)^{f+1}[1+(e/b)\delta]^{a+f+1}\right\}^{-1}$$

La moda de $p(\delta)$ está en $bf[e(a+1)]^{-1}$, la cual debe ser pequeña para asegurar que con gran probabilidad x_e es 'outlier'. Hasta aquí, hemos ignorado la restricción $0 < \delta < 1$. Puede incluirse y usarse integración numérica para determinar las constantes; sin embargo, si la probabilidad inicial para $\delta > 1$ es suficientemente pequeña entonces no se afectan seriamente los resultados. Esta probabilidad segun Pettit(1988), está dada por $I_p(a, f+1)$ donde

$$I_p(g,h) = \int_0^p \frac{u^{g-1}(1-u)^{h-1}}{B(g,h)}du$$

es la función Beta incompleta y $p = be^{-1}/(be^{-1}+1)$. Si be^{-1} es pequeño, entonces la probabilidad será pequeña.

Procediendo con el método tenemos, bajo H_0

$$f(\lambda,\delta)=\frac{b^a e^{f+1}}{\Gamma(a)\Gamma(f+1)}\exp[-b\lambda-e\lambda\delta+a\log\lambda+f\log(\lambda\delta)]$$

$$f(\underline{x}\mid H_0,\lambda,\delta)=\exp\left(-\lambda\sum_{i=1}^{n}x_i+n\log\lambda\right)$$

Con lo que $f(\lambda,\delta\mid \underline{x},H_0)=$

$$\frac{(\sum_{i=1}^{n}x_i+b)^{n+a}e^{f+1}}{\Gamma(n+a)\Gamma(f+1)}\exp\left[-\left(\sum_{i=1}^{n}x_i+b\right)\lambda-e\lambda\delta+(n+a)\log\lambda+f\log(\lambda\delta)\right]$$

De este modo:

$$\begin{aligned}f(\underline{x}\mid H_0)&=\frac{f(\lambda,\delta)f(\underline{x}\mid H_0,\lambda,\delta)}{f(\lambda,\delta\mid H_0,\underline{x})}\\&=\frac{b^a\Gamma(n+a)}{\Gamma(a)(\sum_{i=1}^{n}x_i+b)^{n+a}}\end{aligned}$$

Bajo H_1 tenemos

$$f(\lambda,\delta)=\frac{b^a e^{f+1}}{\Gamma(a)\Gamma(f+1)}\exp[-b\lambda-e\lambda\delta+a\log\lambda+f\log(\lambda\delta)]$$

$$f(\underline{x}\mid H_1,\lambda,\delta)=\exp\left[-\lambda\sum_{i=1}^{n-1}x_i-\lambda\delta x_n+(n-1)\log\lambda+\log(\lambda\delta)\right]$$

Así, $f(\lambda,\delta\mid \underline{x},H_1)=$

$$U\exp\left[-\left(\sum_{i=1}^{n-1}x_i+b\right)\lambda-(e+x_n)\lambda\delta+(n+a-1)\log\lambda+(f+1)\log(\lambda\delta)\right]$$

Donde

$$U=\frac{\left(\sum_{i=1}^{n-1}x_i+b\right)^{n+a-1}(e+x_n)^{f+2}}{\Gamma(n+a-1)\Gamma(f+2)}$$

Con lo que

$$\begin{aligned}f(\underline{x}\mid H_1)&=\frac{f(\lambda,\delta)f(\underline{x}\mid H_1,\lambda,\delta)}{f(\lambda,\delta\mid H_1,\underline{x})}\\&=\frac{(f+1)b^a e^{f+1}\Gamma(n+a-1)}{\Gamma(a)\left(\sum_{i=1}^{n-1}x_i+b\right)^{n+a-1}(e+x_n)^{f+2}}\end{aligned}$$

Siendo el Factor de Bayes que compara el modelo de no 'outliers' contra el de un 'outlier' igual a

$$\begin{aligned} Factor &= \frac{f(\underline{x} \mid H_0)}{f(\underline{x} \mid H_1)} \\ &= \frac{a+n-1}{(f+1)e^{f+1}} \frac{\left(\sum_{i=1}^{n-1} x_i + b\right)^{n+a-1} (e+x_n)^{f+2}}{\left(\sum_{i=1}^{n} x_i + b\right)^{n+a}} \end{aligned}$$

Que podemos expresarlo como

$$\frac{a+n-1}{(f+1)e^{f+1}} \left(1 - \frac{x_n}{\sum_{i=1}^{n} x_i + b}\right)^{n+a-1} \frac{(e+x_n)^{f+2}}{\left(\sum_{i=1}^{n} x_i + b\right)} \tag{5}$$

De nuevo observamos que depende de los datos solamente a través de x_n y de $\sum_{i=1}^{n} x_i$.

Ejemplo 4

La siguiente tabla (Barnett y Lewis, 1978, p.80) contiene una muestra de 131 excesos en el ciclo de tiempo de un proceso de manufactura.

E.c.t. X	Frecuencia
1	18
2	12
3	18
4	16
5	10
6	4
7	9
8	9
9	2
10	7
11	6
12	7
13	2
14	1
15	3
21	3
32	2
35	1
92	1

Como mencionan Barnett y Lewis (1978) es razonable asumir una distribución Exponencial para los datos. La media de los datos es $\overline{x} = 6.44$ y el valor 92 es declarado como 'outlier'.

La distribución conjunta inicial para λ y δ se debe determinar sin tomar en cuenta la muestra, pues ésta debe reflejar el conocimiento previo que se tiene de la muestra, acerca del fenómeno en estudio.

En este caso, para ejemplificar suponemos que antes de tener la muestra se había determinado para λ una distribución marginal inicial como una Gama con parámetros $a = 1.55$ y $b = 10$. Se escogieron dichos valores de los parámetros para que así la media $1/\lambda$ sea aproximadamente igual a 6.44. (suponemos que no sabíamos de antemano que $\bar{x} = 6.44$). El valor de b se escogió igual a 10 para tener poca varianza en la distribución marginal inicial de λ. Para escoger los otros parámetros observamos que también necesitamos asegurarnos de que be^{-1} y $b/[e(a+1)]^{-1}$ sean pequeños. Escogiendo $e = 100$ y $f = 1$ tenemos que la moda para δ es menor que .04 lo cual nos asegura (ver la tabla 1 de la sección anterior) con gran probabilidad, que una observación aberrante será la más grande en la muestra.

Utilizando el Factor de Bayes para comparar el modelo de no 'outliers' contra el de un 'outlier' cuando δ es desconocida (ecuación (5)) tenemos:

$$Factor = .00008$$

que por ser cercano a cero nos indica que, en efecto, el valor 92 es un 'outlier'.

Ejemplo 5

Utilizando los mismos datos del ejemplo anterior, pero ahora se supondrá una distribución inicial de referencia para λ y δ, es decir haciendo $a \longrightarrow 0$, $b \longrightarrow 0$, $e \longrightarrow 1$ y $f \longrightarrow 0$, para reflejar que no tenemos ningún conocimiento inicial hacerca de dichos parámetros, tenemos que el Factor de Bayes para comparar el modelo de no 'outliers' contra el de un 'outlier' utilizando la ecuación (5), tenemos:

$$Factor = .001989$$

que por ser un valor cerca del cero concluimos, al igual que en el ejemplo anterior, que el valor 92 es un 'outlier'.

5.1.1) Modelo Normal$(0,\theta)$

Asumimos que en una muestra de tamaño n, la mayoría de las observaciones tienen una distribución Normal con función de densidad $f(x \mid \theta) = (\frac{\theta}{2\pi})^{1/2}\exp(-\frac{1}{2}\theta x^2)$; $-\infty < x < \infty$. Donde el parámetro θ es la precisión ($\theta = 1/\sigma^2$, con σ^2 la varianza). Consideremos la distribución inicial de θ como una Gama con parámetros α y β (Familia Conjugada). Utilizando el Teorema de la sección **4.4** del Capítulo 4, sabemos que la mínima O.P. para esta distribución está en la más pequeña ó en la más grande observación . A diferencia de la distribución Exponencial, pueden ser detectados 'outliers' superiores y 'outliers' inferiores. De este modo, asumimos que sólo unas pocas observaciones aberrantes pueden tener una distribución Normal con función de densidad

$$g_\delta(x \mid \theta) = \left(\frac{\delta\theta}{2\pi}\right)^{1/2} \exp(-\frac{1}{2}\delta\theta x^2) \quad -\infty < x < \infty,\ 0 < \delta < 1$$

Haremos el desarrollo suponiendo primero que δ es conocida y posteriormente la supondremos desconocida.

En este caso, como el modelo que genera observaciones aberrantes es descrito por una disminución en la precisión (o un incremento en la varianza) del modelo que genera 'buenas' observaciones, para tener que una observación aberrante sea un 'outlier' cambiamos en el modelo básico (sección **4.2** Capítulo 4) el evento

$L_i \equiv$ 'La observación x_i es la más grande en la muestra'.
Por el evento

$L_i \equiv$ 'La observación x_i es la más grande ó la más pequeña en la muestra'.

5.2.2) δ Conocida

Hecha la aclaración anterior, antes que nada, necesitamos ver que se satisfaga (ver Capítulo 4) la aproximación del inciso (2) sección **4.2**. Es decir, debemos tener que con probabilidad alta una observación aberrante será una observación extrema en la muestra. Adecuando algunas de las condiciones dadas por Pettit y Smith (1985) tenemos que una forma de checar que dicha aproximación se cumpla, es encontrar la probabilidad de que una observación aberrante exceda el valor esperado de la estadística de orden max x_i o que sea menor que el mín x_i, $i = 1,\ldots,n-1$, es decir, calcular $p(x_n < \mathrm{E}(\text{mín}x_i)) + p(x_n > \mathrm{E}(\text{max}x_i))$, donde $x_i \sim N(0,\theta)$, $i = 1,\ldots,n-1$ y

$x_n \sim N(0, \delta\theta)$, con $0 < \delta < 1$. En Neave (1928, tabla 2.4(a)) encontramos los valores de dichas esperanzas. Buscando las probabilidades correspondientes hemos construido la siguiente tabla para distintos valores de n y δ.

Tabla 2

Valores de $p(x_n < E(mínx_i)) + p(x_n > E(maxx_i))$

δ \ n	10	15	20	30	50
.11	0.620	0.568	0.526	0.498	0.454
.04	0.764	0.734	0.710	0.684	0.654
.01	0.880	0.864	0.852	0.840	0.820
.0044	0.921	0.910	0.902	0.893	0.881

Construiremos la Estadística de prueba para un 'outlier' superior. Para un 'outlier' inferior es similar.

Para esto necesitamos que la mínima O.P. esté en la observación más grande [1].

Por conveniencia se asumirá que los valores de la muestra son $x_1, \ldots, x_n$ donde x_n es la observación más grande.

Denotando por

$H_0 \equiv$ 'No hay 'outliers' en la muestra'

tenemos

$$f(\theta) \propto \theta^{\alpha-1} \exp(-\beta\theta)$$

$$f(\underline{x} \mid H_0, \theta) = \frac{\theta}{2\pi}^{n/2} \exp(-\frac{1}{2}\theta \sum_{i=1}^{n} x_i^2)$$

con lo que

$$\begin{aligned} f(\theta \mid \underline{x}, H_0) &\propto f(\theta) f(\underline{x} \mid H_0, \theta) \\ &\propto \theta^{n/2+\alpha-1} \exp\left[-\theta\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \beta\right)\right] \end{aligned}$$

introduciendo la constante de proporcionalidad

$$f(\theta \mid \underline{x}, H_0) = \frac{\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \beta\right)^{n/2+\alpha}}{\Gamma(n/2+\alpha)} \theta^{n/2+\alpha-1} \exp\left[-\theta\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \beta\right)\right]$$

---

[1] Para saber si la O.P. toma su valor mínimo en la más grande ó en la más pequeña observación, se comparan $p(x_1 \mid \hat{\theta}_{1n})$ y $p(x_n \mid \hat{\theta}_{1n})$ donde $\hat{\theta}_{1n}$ es el único valor tal que $p(\hat{\theta}_{1n} \mid x_{(1)}) = p(\hat{\theta}_{1n} \mid x_{(n)})$. Después utilizando el Lema de la sección 4.4 podemos saber si $p(x_1 \mid x_{(1)}) < p(x_n \mid x_{(n)})$ ó si $p(x_1 \mid x_{(1)}) > p(x_n \mid x_{(n)})$.

Así, la distribución posterior de θ dado que en la muestra no hay 'outliers' es una Gama con parámetros $n/2+\alpha$ y $\frac{1}{2}\sum_{i=1}^{n} x_i^2+\beta$.

De igual forma, denotando

$H_1 \equiv$ 'Hay un 'outlier' en la muestra'

tenemos

$$f(\theta) \propto \theta^{\alpha-1}\exp(-\beta\theta)$$

$$f(\underline{x} \mid H_1,\theta) = {\frac{\theta}{2\pi}}^{n/2}\delta^{1/2}\exp\left[-\frac{1}{2}\theta\left(\sum_{i=1}^{n-1} x_i^2+\delta x_n^2\right)\right]$$

Con lo cuál

$$f(\theta \mid \underline{x}, H_1) \propto \theta^{n/2+\alpha-1}\delta^{1/2}\exp\left[-\theta\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2+\frac{1}{2}\delta x_n^2+\beta\right)\right]$$

introduciendo la constante de proporcionalidad

$$f(\theta \mid \underline{x}, H_1) = W\exp\left[-\theta\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2+\frac{1}{2}\delta x_n^2+\beta\right)\right]$$

donde

$$W = \frac{\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2+\frac{1}{2}\delta x_n^2+\beta\right)^{n/2+\alpha}}{\Gamma(n/2+\alpha)}\theta^{n/2+\alpha-1}$$

Por lo que la distribución posterior de θ dado que en la muestra hay un 'outlier', es un Gama con parámetros $n/2+\alpha$ y $\frac{1}{2}\sum_{i=1}^{n-1} x_i^2+\frac{1}{2}\delta x_n^2+\beta$.

La probabilidad de que la muestra no contenga 'outliers' es

$$\begin{aligned} f(H_0,\underline{x}) &= f(\underline{x} \mid H_0)f(H_0) \\ &= \left[\Gamma(n/2+\alpha)\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2+\beta\right)^{-(n/2+\alpha)}\right]f(H_0) \end{aligned}$$

La probabilidad de que la muestra contenga un 'outlier' es

$$\begin{aligned} f(H_1,\underline{x}) &= f(\underline{x} \mid H_1)f(H_1) \\ &= \left[\Gamma(n/2+\alpha)\delta^{1/2}\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2+\frac{1}{2}\delta x_n^2+\beta\right)^{-(n/2+\alpha)}\right]f(H_1) \end{aligned}$$

Usando el Factor de Bayes para comparar el modelo en el cuál la muestra no contiene 'outliers' contra el de que contiene un 'outlier' es

$$
\begin{aligned}
Factor &= \frac{f(\underline{x} \mid H_0)}{f(\underline{x} \mid H_1)} \\
&= \frac{\Gamma(n/2+\alpha)\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \beta\right)^{-(n/2+\alpha)}}{\Gamma(n/2+\alpha)\delta^{1/2}\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2 + \delta x_n^2 + \beta\right)^{-(n/2+\alpha)}} \\
&= \delta^{-1/2}\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \beta\right)^{-(n/2+\alpha)}\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2 + \delta x_n^2 + \beta\right)^{(n/2+\alpha)} \\
&= \delta^{-1/2}\left[1 + \frac{\frac{1}{2}x_n^2(\delta-1)}{\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \beta}\right]^{n/2+\alpha}
\end{aligned}
\tag{6}
$$

Siendo entonces el Factor de Bayes una función de $\frac{1}{2}x_n^2\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \beta\right)^{-1}$.

Ejemplo 1

Generamos una muestra de tamaño 50, en la cuál 49 observaciones provienen de la distribución Normal $(0,\theta)$, con $\theta = 4$, utilizando el paquete *STATGRAPHICS*. Suponiendo $\delta = 1/9$ generamos el 'outlier' a través también de *STATGRAPHICS*, siendo dicha observación igual a .94761.

La muestra de tamaño 50 ya ordenada es

-0.88860	-0.85425	-0.79704	-0.66094	-0.65406
-0.57351	-0.54071	-0.50007	-0.46269	-0.37590
-0.32434	-0.30085	-0.29326	-0.23316	-0.21997
-0.19261	-0.17235	-0.14733	-0.10873	-0.08786
-0.07065	-0.04864	-0.03305	-0.00229	0.00250
0.05074	0.06400	0.13259	0.13990	0.14024
0.15986	0.18919	0.27540	0.28613	0.29632
0.32301	0.37573	0.42309	0.43585	0.44940
0.45637	0.50994	0.52644	0.56759	0.58810
0.59047	0.65448	0.76122	0.85182	0.94761

La distribución incial de θ la suponemos como una Gama con $\alpha = 8$ y $\beta = 2$, para que así la media resulte ser 4. Con los datos anteriores tenemos que el Factor de Bayes que nos compara el modelo de no 'outliers' con el de un 'outlier' (ecuación (6)) resulta

$$Factor = .452$$

Como este valor es menor que uno, nos apoya la hipótesis de que el valor .94761 es un 'outlier' (ver sección **2.6** del Capítulo 2).

Ejemplo 2

Suponiendo la misma muestra, pero ahora utilizando una $\delta = 1/100$ la observación generada a través del *STATGRAPHICS* resulta ser 4.3281. Para estos datos el Factor de Bayes resulta ser

$$Factor = 4.8 \times 10^{-12}$$

Así, como el Factor resulta ser cercano a cero nos está apoyando el hecho de que el valor 4.3281 es un 'outlier'.

En este último ejemplo es "más claro" el apoyo de que el valor más grande sea un 'outlier', ya que el valor de $\delta = 1/100$ nos asegura con gran probabilidad que una observación aberrante será la más grande en la muestra. (Ver tabla 2).

Consideremos ahora la posibilidad de tener dos 'outliers' superiores, x_n y x_{n-1}. Denotamos

$H_2 \equiv$ 'Hay dos 'outliers' en la muestra'

Así

$$f(\theta) \propto \theta^{\alpha-1}\exp(-\beta\theta)$$

$$f(\underline{x} \mid H_2, \theta) = \left(\frac{\theta}{2\pi}\right)^{n/2} \delta \exp\left[-\frac{1}{2}\theta\left(\sum_{i=1}^{n-2} x_i^2 + \delta(x_n^2 + x_{n-1}^2)\right)\right]$$

Por lo que

$$f(\theta \mid \underline{x}, H_2) \propto \theta^{n/2+\alpha-1}\delta \exp\left[-\theta\left(\frac{1}{2}\sum_{i=1}^{n-2} x_i^2 + \frac{1}{2}\delta(x_n^2 + x_{n-1}^2) + \beta\right)\right]$$

introduciendo la constante de proporcionalidad

$$= W \exp\left[-\theta\left(\frac{1}{2}\sum_{i=1}^{n-2} x_i^2 + \frac{1}{2}\delta(x_n^2 + x_{n-1}^2) + \beta\right)\right]$$

donde

$$W = \frac{\left(\frac{1}{2}\sum_{i=1}^{n-2} x_i^2 + \frac{1}{2}\delta(x_n^2 + x_{n-1}^2) + \beta\right)^{n/2+\alpha}}{\Gamma(n/2+\alpha)}\theta^{n/2+\alpha-1}$$

Siendo entonces una Gamma con parámetros $n/2+\alpha$ y $\frac{1}{2}\sum_{i=1}^{n-2} x_i^2 + \frac{1}{2}\delta(x_n^2 + x_{n-1}^2) + \beta$.

Con lo anterior tenemos que

$$f(H_2, \underline{x}) = f(\underline{x} \mid H_2) f(H_2)$$

$$= \left[\Gamma(n/2+\alpha)\delta\left(\frac{1}{2}\sum_{i=1}^{n-2} x_i^2 + \frac{1}{2}\delta(x_n^2 + x_{n-1}^2) + \beta\right)^{-(n/2+\alpha)}\right] f(H_2)$$

Por lo que el Factor de Bayes para comparar el modelo de no 'outliers' con el de dos 'outliers' es

$$
(7)\qquad \begin{aligned} Factor &= \delta^{-1}\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2+\beta\right)^{-(n/2+\alpha)}\left(\frac{1}{2}\sum_{i=1}^{n-2} x_i^2+\frac{1}{2}\delta(x_n^2+x_{n-1}^2)+\beta\right)^{(n/2+\alpha)} \\ &= \delta^{-1}\left[1+\frac{\frac{1}{2}(x_n^2+x_{n-1}^2)}{\frac{1}{2}\sum_{i=1}^{n} x_i^2+\beta}(\delta-1)\right]^{n/2+\alpha} \end{aligned}
$$

Siendo una función de $\frac{1}{2}(x_n^2+x_{n-1}^2)(\frac{1}{2}\sum_{i=1}^{n} x_i^2+\beta)^{-1}$. Esta estadística es suceptible a efectos engañosos. Si x_n es muy grande, entonces el par x_n y x_{n-1} pueden ser delcarados como 'outliers' aunque x_{n-1} no lo sea.

El Factor de Bayes del modelo de un 'outlier' contra el de dos 'outliers' es

$$
(8)\qquad \begin{aligned} Factor &= \delta^{-1/2}\frac{\left(\frac{1}{2}\sum_{i=1}^{n-2} x_i^2+\frac{1}{2}\delta(x_n^2+x_{n-1}^2)+\beta\right)^{(n/2+\alpha)}}{\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2+\frac{1}{2}\delta x_n^2+\beta\right)^{(n/2+\alpha)}} \\ &= \delta^{-1/2}\left[1+\frac{\frac{1}{2}x_{n-1}^2}{\frac{1}{2}\sum_{i=1}^{n-1} x_i^2+\frac{1}{2}\delta x_n^2+\beta}(\delta-1)\right]^{n/2+\alpha} \end{aligned}
$$

Siendo una función de $\frac{1}{2}x_{n-1}^2(\frac{1}{2}x_1^2+\ldots+\frac{1}{2}x_{n-1}^2+\frac{1}{2}\delta x_n^2+\beta)^{-1}$.

Lo cual nos asegura que, si bien un valor grande de x_n puede llevarnos a tener poca probabilidad para no 'outlier' contra dos 'outliers', es necesario un valor grande de x_{n-1} para tener gran peso para dos 'outliers' contra un 'outlier'.

Además, de igual forma podemos tener que si bien un valor grande de x_{n-1} puede causarnos gran peso para no 'outliers' contra un 'outlier', tenemos poco peso para no 'outliers' contra dos 'outliers'.

Ejemplo 3

Suponiendo los datos del ejemplo 2, tenemos que el Factor de Bayes que nos compara el modelo de no 'outliers' con el de dos 'outliers' (ecuación (7)) es

$$Factor = 8.04\times 10^{-12}$$

que por ser cercano a cero nos apoya que los valores $x_{50}=4.3281$ y $x_{49}=.85182$ son ambos 'outliers', aunque éste último no lo sea.

Ahora, utilizando el Factor de Bayes para comparar el modelo de un 'outlier' con el de dos 'outliers' (ecuación (8)), se tiene

$$Factor = 1.672091$$

que por ser mayor que 1 nos indica que sólo hay un 'outlier' en la muestra, es decir, que sólo el valor 4.3281 es 'outlier'.

5.2.3) δ Desconocida

Si δ es desconocida debemos especificar una distribución conjunta inicial para θ y δ. Si se tienen $n-1$ observaciones $x_1,\ldots,x_{n-1}$ de la distribución Normal $f(x \mid \theta) = \left(\frac{\theta}{2\pi}\right)^{1/2}\exp(-\frac{1}{2}\theta x^2)$, $-\infty < x < \infty$ y una observación, x_n, de la distribución Normal $g(x \mid \theta) = \left(\frac{\delta\theta}{2\pi}\right)^{1/2}\exp(-\frac{1}{2}\delta\theta x^2)$, entonces la verosimilitud en la forma general de la Familia Exponencial es proporcional a

$$\exp\left[-\frac{1}{2}\theta\sum_{i=1}^{n-1}x_i^2 - \frac{1}{2}\theta\delta x_n^2 + \frac{(n-1)}{2}\log\theta + \frac{1}{2}\log(\theta\delta)\right]$$

lo cual nos sugiere utilizar una conjugada inicial como

$$p(\theta,\delta) \propto exp[-b\theta - e\theta\delta + a\log\theta + f\log(\theta\delta)]$$

donde la constante de proporcionalidad está dada por

$$b^a e^{f+1}\,[\Gamma(f+1)\Gamma(a)]^{-1}$$

Ver sección 5.1.3

La distribución marginal inicial para θ es una Gama con parámetros a y b y para δ es Siegel (ver Apéndice C).

La moda de $p(\delta)$ está en $bf[e(a+1)]^{-1}$, la cuál debe ser pequeña para asegurar que con gran probabilidad x_n es 'outlier'. Hasta aquí, se ha ignorado la restricción de $0 < \delta < 1$. Puede incluirse y usarse integración numérica para determinar las constantes; sin embargo, si la probabilidad inicial para $\delta > 1$ es suficientemente pequeña entonces no se afectan seriamente los resultados. Esta probabilidad segun Pettit(1983), está dada por $I_p(a, f+1)$ donde

$$I_p(g,h) = \int_0^p \frac{u^{g-1}(1-u)^{h-1}}{B(g,h)}du$$

es la función Beta incompleta y $p = be^{-1}/(be^{-1}+1)$. Si be^{-1} es pequeño entonces la probabilidad será pequeña.

Procediendo con el método tenemos, bajo $\bar{H}_0$

$$f(\theta,\delta) = \frac{b^a e^{f+1}}{\Gamma(a)\Gamma(f+1)}\exp[-b\theta - e\theta\delta + a\log\theta + f\log(\theta\delta)]$$

$$f(\underline{x} \mid H_0,\theta,\delta) \propto \exp\left(-\frac{1}{2}\theta\sum_{i=1}^{n}x_i^2 + \frac{n}{2}\log\theta\right)$$

Con lo que $f(\theta,\delta \mid \underline{x}, H_0) =$

$$V\exp\left[-\left(\frac{1}{2}\sum_{i=1}^{n}x_i^2 + b\right)\theta - e\theta\delta + (n/2+a)\log\theta + f\log(\theta\delta)\right]$$

Donde

$$V = \frac{\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + b\right)^{n/2+a} e^{f+1}}{\Gamma(n/2+a)\Gamma(f+1)}$$

De este modo:

$$f(\underline{x} \mid H_0) = \frac{b^a \Gamma(n/2+a)}{\Gamma(a)\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + b\right)^{n/2+a}}$$

Bajo H_1 tenemos

$$f(\theta,\delta) = \frac{b^a e^{f+1}}{\Gamma(a)\Gamma(f+1)} \exp[-b\theta - e\theta\delta + a\log\theta + f\log(\theta\delta)]$$

$$f(\underline{x} \mid H_1,\theta,\delta) = \exp\left[-\frac{1}{2}\theta\sum_{i=1}^{n-1} x_i^2 - \frac{1}{2}\theta\delta x_n^2 + \frac{(n-1)}{2}\log\theta + \frac{1}{2}\log(\theta\delta)\right]$$

Así, $f(\theta,\delta \mid \underline{x}, H_1) =$

$$U \exp\left[-\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2 + b\right)\theta - (e + \frac{1}{2}x_n^2)\theta\delta + (n/2 + a - 1/2)\log\theta + (f + 1/2)\log(\theta\delta)\right]$$

Donde

$$U = \frac{\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2 + b\right)^{n/2+a-1/2}(e + \frac{1}{2}x_n^2)^{f+3/2}}{\Gamma(n/2+a-1/2)\Gamma(f+3/2)}$$

Con lo cual

$$f(\underline{x} \mid H_1) = \left(\frac{\Gamma(f+3/2)}{\Gamma(f+1)}\right)\frac{b^a e^{f+1}\Gamma(n/2+a-1/2)}{\Gamma(a)\left(\frac{1}{2}\sum_{i=1}^{n-1} x_i^2 + b\right)^{n/2+a-1/2}(e + \frac{1}{2}x_n^2)^{f+3/2}}$$

Siendo entonces el Factor de Bayes que compara el modelo de no 'outliers' contra el de un 'outlier' igual a

$$Factor = \frac{f(\underline{x} \mid H_0)}{f(\underline{x} \mid H_1)}$$

$$= \left(\frac{\Gamma(f+1)\Gamma(n/2+a)}{\Gamma(f+3/2)\Gamma(n/2+a-1/2)e^{f+1}}\right)\frac{\left(\sum_{i=1}^{n-1} x_i^2 + b\right)^{n/2+a-1/2}(e + x_n^2)^{f+2}}{\left(\sum_{i=1}^{n} x_i^2 + b\right)^{n/2+a}}$$

Que podemos expresarlo como

(9)

$$\frac{\Gamma(f+1)\Gamma(n/2+a)}{\Gamma(f+3/2)\Gamma(n/2+a-1/2)e^{f+1}}\left(1 - \frac{\frac{1}{2}x_n^2}{\frac{1}{2}\sum_{i=1}^{n} x_i^2 + b}\right)^{n/2+a-1/2}\frac{(e + \frac{1}{2}x_n^2)^{f+3/2}}{\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 + b\right)^{1/2}}$$

Que depende de los datos solamente a través de x_k^2 y de $\sum_{i=1}^{n} x_i^2$.

Ejemplo 4

Los siguientes datos tomados de Rosner (1983), corresponden a una muestra de 52 observaciones que se suponen de una distribución Normal con media cero y precisión desconocida.

-2.377	-1.447	-1.187	-0.977	-0.927
-0.867	-0.867	-0.787	-0.747	-0.697
-0.637	-0.637	-0.577	-0.567	-0.547
-0.477	-0.437	-0.427	-0.367	-0.357
-0.317	-0.217	-0.187	-0.167	-0.137
-0.067	-0.037	-0.027	0.013	0.023
0.103	0.113	0.133	0.223	0.243
0.273	0.343	0.413	0.493	0.513
0.773	0.793	0.793	0.793	1.083
1.083	1.173	1.463	1.553	2.173
2.513	3.883			

El valor 3.883 es declarado como un 'outlier' por Rosner. Tenemos que la varianza muestral, sin tomar en cuenta la última observación es igual a .7962, siendo entonces la precisión igual a 1.256. Asumimos que θ tiene una distribución marginal inicial Gama con parámetros $a = 25.12$ y $b = 0.05$. [2].

Como además necesitamos que be^{-1} y $b/[e(a+1)]^{-1}$ sean pequeños escogemos $e = 3$ y $f = 3$; teniendo con estos valores que la moda para δ es menor que .002 lo cuál por la tabla de la del principio de la sección 5.2.2 nos asegura que con probabilidad alta una observación aberrante será la más grande en la muestra.

Utilizando el Factor de Bayes para comparar el modelo de no 'outliers' contra el de un 'outlier' cuando θ es desconocida (ecuación (9)) tenemos:

$$Factor = 3.2333 \times 10^{-5}$$

lo cuál nos indica, por ser un valor cerca del cero, que, en efecto, el valor 3.883 es un 'outlier'.

Ejemplo 5

Utilizando los mismos datos del ejemplo anterior, pero ahora se supondrá una distribución inicial de referencia para θ y δ, es decir haciendo $a \longrightarrow 0$, $b \longrightarrow 0$, $e \longrightarrow 1$ y $f \longrightarrow 0$, para reflejar que no tenemos ningún conocimiento inicial hacerca de dichos

[2] La distribución conjunta inicial para θ y δ se debe determinar sin tomar en cuenta la muestra, pues ésta debe reflejar el conocimiento que se tiene previo a la muestra, acerca del fenómeno en estudio.

parámetros, tenemos que el Factor de Bayes para comparar el modelo de no 'outliers' contra el de un 'outlier' utilizando la ecuación (9), tenemos:

$$Factor = 7.5011 \times 10^{-3}$$

que por ser un valor cerca del cero conlcuimos, al igual que en ejemplo anterior, que el valor 3.883 es un 'outlier'.

5.3 Distribución Normal$(\theta,1)$

5.3.1) Modelo Normal$(\theta,1)$

Asumimos que en una muestra de tamaño n, la mayoría de las observaciones tienen una distribución Normal con función de densidad $f(x \mid \theta) = (\frac{1}{2\pi})^{1/2} \exp\left(-\frac{1}{2}(x-\theta)^2\right)$; $-\infty < x < \infty$. Donde el parámetro θ es la media de la distribución. Consideremos la distribución inicial de θ como una Normal con parámetros α y β (Familia Conjugada). Utilizando el Teorema de la sección **4.4** del Capítulo 4, sabemos que la mínima Ordenada Predictiva para esta distribución está en la más pequeña ó en la más grande observación. En esta distribución pueden ser detectados 'outliers' superiores y 'outliers' inferiores. Supondremos que la Ordenada Predictiva toma su mínimo valor en la más grande de las observaciones [3] y desarrollaremos el modelo para 'outliers' superiores. Para los inferiores es similar.

Así, asumimos que sólo unas pocas observaciones aberrantes pueden tener una distribución Normal con función de densidad

$$g_\delta(x \mid \theta) = \left(\frac{1}{2\pi}\right)^{1/2} \exp\left(-\frac{1}{2}(x-(\theta+\delta))^2\right) \quad -\infty < x < \infty,\ \delta > 0$$

Haremos el desarrollo suponiendo primero que δ es conocida y posteriormente la supondremos desconocida.

5.3.2) δ Conocida

Antes que nada, necesitamos ver que se satisfaga la aproximación del inciso (2) sección **4.2**. Es decir, debe cumplirse que con probabilidad alta una observación aberrante sea la más grande en la muestra. Adecuando algunas de las condiciones dadas por Pettit y Smith (1985) tenemos que una forma de checar que dicha aproximación se cumpla, es encontrar la probabilidad de que una observación aberrante exceda el valor esperado de la estadística de orden max x_i, es decir, calcular $p(x_n > \mathrm{E}(\max x_i))$, donde $x_i \sim N(\theta,1)$, $i = 1,\ldots,n-1$ y $x_n \sim N(\theta+\delta,1)$, con $\delta > 0$. En Neave (1928, tabla 2.4(a)) encontramos los valores de dichas esperanzas. Buscando las probabilidades correspondientes se ha construido la siguiente tabla para distintos valores de δ, cuando la precisión es igual a 1.

---

[3] Para saber si la Ordenada Predictiva toma su valor mínimo en la más grande ó en la más pequeña observación, se comparan $p(x_1 \mid \hat{\theta}_{1n})$ y $p(x_n \mid \hat{\theta}_{1n})$ donde $\hat{\theta}_{1n}$ es el único valor tal que $p(\hat{\theta}_{1n} \mid x_{(1)} = p(\hat{\theta}_{1n} \mid x_{(n)}$. Después utilizando el Lema de la sección **4.4** podemos saber si $p(x_1 \mid x_{(1)}) < p(x_n \mid x_{(n)})$ ó si $p(x_1 \mid x_{(1)}) > p(x_n \mid x_{(n)})$.

Tabla 3

Valores de $p(x_n > E(\max x_i))$

$\delta \setminus n$	10	15	20	30	50
3	0.935	0.903	0.876	0.834	0.776
3.5	0.978	0.964	0.951	0.929	0.896
4	0.994	0.989	0.984	0.976	0.960
5	0.999	0.999	0.999	0.998	0.997

Por conveniencia se asumirá que los valores de la muestra son $x_1, \ldots, x_n$ donde x_n es la observación más grande.

Denotando por

$H_0 \equiv$ 'No hay 'outliers' en la muestra'

tenemos

$$f(\theta) \propto \exp\left[-\frac{\beta}{2}(\alpha-\theta)^2\right]$$

$$f(\underline{x} \mid H_0, \theta) \propto \exp\left[-\frac{n}{2}(\theta-\overline{x})^2\right]$$

con lo que

$$\begin{aligned} f(\theta \mid \underline{x}, H_0) &\propto f(\theta) f(\underline{x} \mid H_0, \theta) \\ &\propto \exp\left[-\frac{(\beta+n)}{2}(\theta-\mu')^2\right] \end{aligned}$$

donde

$$\mu' = \frac{\beta\alpha + n\overline{x}}{\beta+n}$$

Así, la distribución posterior de θ dado que en la muestra no hay 'outliers' es una Normal con media μ' y precisión $\beta + n$.

No introduciremos las constantes de proporcionalidad pues son factores comunes en todas las distribuciones que obtengamos y no afectan por lo tanto en el Factor de Bayes.

De igual forma, denotando

$H_1 \equiv$ 'Hay un 'outlier' en la muestra'

tenemos

$$f(\theta) \propto \exp\left[-\frac{\beta}{2}(\alpha-\theta)^2\right]$$

$$f(\underline{x} \mid H_1, \theta) \propto \exp\left[-\frac{n}{2}\left(\theta-(\overline{x}-\delta/n)\right)^2\right]$$

con lo cuál

$$f(\theta \mid \underline{x}, H_1) \propto \exp\left[-\frac{(\beta+n)}{2}(\theta-\mu'')^2\right]$$

donde

$$\mu'' = \frac{\beta\alpha + n\overline{x} - \delta}{\beta + n}$$

Por lo que la distribución posterior de θ dado que en la muestra hay un 'outlier', es un Normal con media μ'' y precisión $\beta + n$.

La probabilidad de que la muestra no contenga 'outliers' es

$$\begin{aligned} f(H_0, \underline{x}) &= f(\underline{x} \mid H_0) f(H_0) \\ &= \exp\left[-\frac{n\beta}{2(\beta+n)}(\alpha-\overline{x})^2\right] f(H_0) \end{aligned}$$

La probabilidad de que la muestra contenga un 'outlier' es

$$\begin{aligned} f(H_1, \underline{x}) &= f(\underline{x} \mid H_1) f(H_1) \\ &= \exp\left[-\frac{n\beta}{2(\beta+n)}(\alpha-\overline{x})^2 + \frac{2n\beta\overline{x}\delta - \delta^2\beta - 2n\delta\alpha\beta}{2n(n+\beta)}\right] f(H_1) \end{aligned}$$

Usando el Factor de Bayes para comparar el modelo en el cuál la muestra no contiene 'outliers' contra el de que contiene un 'outlier' es

$$\begin{aligned} Factor &= \frac{f(\underline{x} \mid H_0)}{f(\underline{x} \mid H_1)} \\ &= \exp\left[\frac{-2n\beta\overline{x}\delta + \delta^2\beta + 2n\delta\alpha\beta}{2n(n+\beta)}\right] \quad \delta > 0 \\ &= \exp\left[\frac{\beta\delta(\delta + 2n\alpha - 2n\overline{x})}{2n(n+\beta)}\right] \quad \delta > 0 \end{aligned} \tag{10}$$

Ejemplo 1

Generamos una muestra de tamaño 50, en la cuál 49 observaciones provienen de la distribución Normal $(\theta, 1)$, con $\theta = 10$, utilizando el paquete *STATGRAPHICS*. Suponiendo $\delta = 3$ generamos el 'outlier' a través también de *STATGRAPHICS*, siendo dicha observación igual a 14.56.

La muestra de tamaño 50 ya ordenada es

7.12	7.71	8.28	8.57	8.79
8.82	9.04	9.09	9.10	9.25
9.32	9.40	9.49	9.52	9.54
9.55	9.57	9.58	9.64	9.67
9.75	9.80	9.85	10.02	10.07
10.10	10.12	10.14	10.20	10.20
10.21	10.29	10.36	10.39	10.43
10.60	10.75	10.80	10.94	11.01
11.05	11.06	11.32	11.30	11.36
11.39	11.64	11.86	11.98	14.56

La distribución incial de θ la suponemos como una Normal(α, β), con $\alpha = 10$ y $\beta = 100$. Con los datos anteriores tenemos que el Factor de Bayes que nos compara el modelo de no 'outliers' con el de un 'outlier' (ecuación (10)) resulta

$$Factor = .8833$$

Como este valor es menor que uno, nos apoya la hipótesis de que el valor 14.56 es un 'outlier' (ver sección 2.6 del Capítulo 2). Dicho apoyo no es muy fuerte, observando que el Factor de Bayes resulta menor que .5 cuando el valor del 'outlier' es mayor a 28.25.

Ejemplo 2

Suponiendo la misma muestra, pero ahora utilizando una $\delta = 4.5$ la observación generada a través del *STATGRAPHICS* resulta ser 16.2 . Para estos datos el Factor de Bayes resulta ser

$$Factor = .7879$$

Así, como el Factor resulta ser menor que uno nos está apoyando el hecho de que el valor 16.2 es un 'outlier'.

Al igual que en el ejemplo anteior no es muy fuerte el apoyo de que el valor 16.2 sea un 'outlier', siendo el Factor de Bayes menor que .5 cuando el valor del 'outlier' es mayor a 23.79.

Consideremos ahora la posibilidad de tener dos 'outliers' superiores, x_n y x_{n-1}. Denotamos

$H_2 \equiv$ 'Hay dos 'outliers' en la muestra'

Así

$$f(\theta) \propto \exp\left[-\frac{\beta}{2}(\alpha-\theta)^2\right]$$
$$f(\underline{x} \mid H_2, \theta) \propto \exp\left[-\frac{n}{2}\left(\theta - (\overline{x} - 2\delta/n)\right)^2\right]$$

con lo cuál

$$f(\theta \mid \underline{x}, H_2) \propto \exp\left[-\frac{(\beta+n)}{2}(\theta - \mu''')^2\right]$$

donde

$$\mu''' = \frac{\beta\alpha + n\overline{x} - 2\delta}{\beta + n}$$

siendo entonces una Normal con media μ''' y precisión $\beta + n$.

Con lo anterior tenemos que

$$\begin{aligned} f(H_2, \underline{x}) &= f(\underline{x} \mid H_2) f(H_2) \\ &= \exp\left[-\frac{n\beta}{2(\beta+n)}(\alpha - \overline{x})^2 + \frac{2n\beta\overline{x}\delta - 2\delta^2\beta - 2n\delta\alpha\beta}{n(n+\beta)}\right] f(H_2) \end{aligned}$$

Por lo que el Factor de Bayes para comparar el modelo de no 'outliers' con el de dos 'outliers' es

$$
\text{(11)} \qquad \begin{aligned} Factor &= \exp\left[\frac{-4n\beta\bar{x}\delta + 4\delta^2\beta + 4n\delta\alpha\beta}{2n(n+\beta)}\right] \quad \delta > 0 \\ &= \exp\left[\frac{\beta\delta(4\delta + 4n\alpha - 4n\bar{x})}{2n(n+\beta)}\right] \quad \delta > 0 \end{aligned}
$$

El Factor de Bayes del modelo de un 'outlier' contra el de dos 'outliers' es

$$
\text{(12)} \qquad \begin{aligned} Factor &= \exp\left[\frac{-2n\beta\bar{x}\delta + 3\delta^2\beta + 2n\delta\alpha\beta}{2n(n+\beta)}\right] \quad \delta > 0 \\ &= \exp\left[\frac{\beta\delta(3\delta + 2n\alpha - 2n\bar{x})}{2n(n+\beta)}\right] \quad \delta > 0 \end{aligned}
$$

Ejemplo 3

En los datos del ejemplo 2, supongamos que se tienen los 'outliers' $x_{49} = 15.6$ y $x_{50} = 16.2$. Entonces el Factor de Bayes que nos compara el modelo de no 'outliers' con el de dos 'outliers' (ecuación (11)) es

$$
Factor = .5256
$$

que por ser menor a 1 nos apoya que los valores $x_{50} = 16.2$ y $x_{49} = 15.6$ son ambos 'outliers'.

5.2.3) δ Desconocida

Como δ es desconocida debemos especificar una distribución conjunta inicial para θ y δ. Si se tienen $n-1$ observaciones $x_1, \ldots, x_{n-1}$ de la distribución Normal $(\theta, 1)$ y una observación, x_n, de la distribución Normal $(\theta+\delta, 1)$ $\delta > 0$ entonces la verosimilitud en la forma general de la Familia Exponencial es proporcional a

$$
\exp\left[-\frac{n}{2}\theta^2 + n\bar{x}\theta - \theta\delta - \frac{1}{2n}\delta^2 + \bar{x}\delta\right]
$$

lo cual nos sugiere utilizar una conjugada inicial como

$$
p(\theta, \delta) \propto \exp\left[-a\theta^2 + b\theta - c\theta\delta - e\delta^2 + f\delta\right]
$$

donde la constante de proporcionalidad está dada por el inverso multiplicativo de:

$$
\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-a\theta^2 + b\theta - c\theta\delta - e\delta^2 + f\delta\right] d\delta d\theta \\ &= \int_{-\infty}^{\infty} \exp[-a\theta^2 + b\theta]\left[\int_{-\infty}^{\infty} \exp[-e\delta^2 + f\delta - c\theta\delta] d\delta\right] d\theta \\ &= \int_{-\infty}^{\infty} \exp[-a\theta^2 + b\theta] \exp\left[\frac{(f-c\theta)^2}{4e}\right]\left(\frac{e}{\pi}\right)^{-1/2}\left[\int_{-\infty}^{\infty} A\, d\delta\right] d\theta \end{aligned}
$$

donde

$$A=\left(\frac{e}{\pi}\right)^{1/2}\exp\left[-e\left(\delta-\frac{f-c\theta}{2e}\right)^2\right]$$

La integral respecto a δ, es la integral de una Normal con media $(f-c\theta)/2e$ y precisión $2e$, por tanto es igual a 1.

$$=\left(\frac{e}{\pi}\right)^{-1/2}\exp\left(\frac{f^2}{e4}\right)\int_{-\infty}^{\infty}\exp\left[-\left(\frac{4ae-c^2}{4e}\right)\theta^2+\left(\frac{2be-fc}{2e}\right)\theta\right]d\theta$$
$$=W\int_{-\infty}^{\infty}\left(\frac{4ae-c^2}{4e\pi}\right)^{1/2}\exp\left[-\left(\frac{4ae-c^2}{4e}\right)\left(\theta-\frac{1}{2}\left(\frac{4be-2fc}{4ae-c^2}\right)\right)^2\right]d\theta$$

donde

$$W=\left(\frac{e}{\pi}\right)^{-1/2}\exp\left(\frac{f^2}{e4}\right)\left(\frac{4ae-c^2}{4e\pi}\right)^{-1/2}\exp\left[\frac{1}{16e}\frac{(4be-2fc)^2}{4ae-c^2}\right]$$

La integral respecto a θ vale 1 ya que es la integral de una Normal con media $(4be-2fc)/2(4ae-c^2)$ y precisión $(4ae-c^2)/2e$.

De esta forma la constante de proporcionalidad es igual a $1/W$, es decir,

$$K=1/W=\frac{(4ae-c^2)^{1/2}}{2\pi}\exp\left[-\frac{af^2+eb^2-bcf}{4ae-c^2}\right] \tag{13}$$

La distribución marginal inicial para θ es

$$\int_{-\infty}^{\infty}\frac{(4ae-c^2)^{1/2}}{2\pi}\exp\left[-\frac{af^2+eb^2-bcf}{4ae-c^2}\right]\exp\left[-a\theta^2+b\theta-c\theta\delta-e\delta^2+f\delta\right]d\delta$$
$$=\left(\frac{4ae-c^2}{4e\pi}\right)^{1/2}\exp\left[\frac{-(4ae-c^2)}{4e}\left(\theta-\frac{2be-fc}{4ae-c^2}\right)^2\right]$$

la cuál es una Normal con media $(2be-fc)/(4ae-c^2)$ y precisión $(4ae-c^2)/2e$.

De igual forma la distribución marginal inicial para δ es

$$\int_{-\infty}^{\infty}\frac{(4ae-c^2)^{1/2}}{2\pi}\exp\left[-\frac{af^2+eb^2-bcf}{4ae-c^2}\right]\exp\left[-a\theta^2+b\theta-c\theta\delta-e\delta^2+f\delta\right]d\theta$$
$$=\left(\frac{4ae-c^2}{4a\pi}\right)^{1/2}\exp\left[\frac{-(4ae-c^2)}{4a}\left(\delta-\frac{2af-bc}{4ae-c^2}\right)^2\right]$$

La cuál es una Normal con media $(2af-bc)/(4ae-c^2)$ y precisión $(4ae-c^2)/2a$.

La moda de $p(\delta)$ es la media de la distribución la cuál debe ser mayor que cero asegurar que x_n es 'outlier' superior. Hasta aquí, se ha ignorado la restricción de $\delta>0$. Puede incluirse y usarse integración numérica para determinar las constantes. Pero si controlamos que la probabilidad inicial para $\delta < 0$ sea suficientemente pequeña entonces no se afectan seriamente los resultados. Esta probabilidad está dada por

$$\int_{-\infty}^{0}\left(\frac{4ae-c^2}{4a\pi}\right)^{1/2}\exp\left[\frac{-(4ae-c^2)}{4a}\left(\delta-\frac{2af-bc}{4ae-c^2}\right)^2\right]d\delta$$
$$=\int_{-\infty}^{c}\left(\frac{1}{2\pi}\right)^{1/2}\exp\left[\frac{-1}{2}Z^2\right]dZ$$

donde

$$c=-\frac{2af-bc}{[2a(4ae-c^2)]^{1/2}}$$

Por tanto $(2af-bc)[2a(4ae-c^2)]^{-1/2}$ debe ser grande, en particular mayor que cero por lo descrito anteriormente, para que la probabilidad sea pequeña.

Procediendo con el método tenemos, bajo H_0

$$f(\theta,\delta)=K\exp[-a\theta^2+b\theta-c\theta\delta-e\delta^2+f\delta]$$

Con K, como en la ecuación (13).

$$f(\bar{x}\mid H_0,\theta,\delta)=\left(\frac{n}{2\pi}\right)^{1/2}\exp\left[-\frac{n}{2}(\theta-\bar{x})^2\right]$$

con lo que $f(\theta,\delta\mid\bar{x},H_0)=$

$$W\exp[-(a+n/2)\theta^2+(b+n\bar{x})\theta-c\theta\delta-e\delta^2+f\delta]$$

donde

$$W=\frac{[4(a+n/2)e-c^2]^{1/2}}{2\pi}\exp\left[\frac{-(a+n/2)f^2-e(b+n\bar{x})^2+(b+n\bar{x})cf}{4(a+n/2)e-c^2}\right]$$

De este modo:

$$f(\bar{x}\mid H_0)=M\left(\frac{4ae-c^2}{4(a+n/2)e-c^2}\right)^{1/2}\exp\left[\frac{(a+n/2)f^2+e(b+n\bar{x})^2-(b+n\bar{x})cf}{4(a+n/2)e-c^2}\right]$$

con

$$M=\left(\frac{n}{2\pi}\right)^{1/2}\exp\left(\frac{-n\bar{x}^2}{2}\right)\exp\left[\frac{-af^2-eb^2+bcf}{4ae-c^2}\right]\tag{14}$$

Bajo H_1 tenemos

$$f(\theta,\delta)=K\exp[-a\theta^2+b\theta-c\theta\delta-e\delta^2+f\delta]$$

$$f(\bar{x}\mid H_1,\theta,\delta)=\left(\frac{n}{2\pi}\right)^{1/2}\exp\left[-\frac{n}{2}(\theta-(\bar{x}-\delta/n))^2\right]$$

Con lo que $f(\theta, \delta \mid \underline{x}, H_1) =$

$$W \exp[-(a+n/2)\theta^2 + (b+n\bar{x})\theta - (c+1)\theta\delta - (e+\frac{1}{2n}\delta^2 + (f+\bar{x})\delta)]$$

donde

$$W = \frac{[4(a+n/2)(e+\frac{1}{2n}) - (c+1)^2]^{1/2}}{2\pi} \exp[w]$$

con

$$w = \frac{-(a+n/2)(f+\bar{x})^2 - (e+\frac{1}{2n})(b+n\bar{x})^2 + (b+n\bar{x})(c+1)(f+\bar{x})}{4(a+n/2)(e+\frac{1}{2n}) - (c+1)^2}$$

Con lo cual

$$f(\underline{x} \mid H_1) = M\left(\frac{4ae - c^2}{4(a+n/2)(e+\frac{1}{2n}) - (c+1)^2}\right)^{1/2} \exp[m]$$

Con M de la forma de la expresión (14) y

$$m = \frac{(a+n/2)(f+\bar{x})^2 + (e+\frac{1}{2n})(b+n\bar{x})^2 - (b+n\bar{x})(c+1)(f+\bar{x})}{4(a+n/2)(e+\frac{1}{2n}) - (c+1)^2}$$

Siendo entonces el Factor de Bayes que compara el modelo de no 'outliers' contra el de un 'outlier' igual a

(15)
$$\begin{aligned} Factor &= \frac{f(\underline{x} \mid H_0)}{f(\underline{x} \mid H_1)} \\ &= (F_1)^{1/2} \exp[F_2] \exp[F_3] \end{aligned}$$

con

$$F_1 = \frac{4(a+n/2)(e+\frac{1}{2n}) - (c+1)^2}{4(a+n/2)e - c^2}$$

$$F_2 = \exp\left[\frac{(a+n/2)f^2 + e(b+n\bar{x})^2 - (b+n\bar{x})cf}{4(a+n/2)e - c^2}\right]$$

$$F_3 = \exp\left[\frac{-(a+n/2)(f+\bar{x})^2 - (e+\frac{1}{2n})(b+n\bar{x})^2 + (b+n\bar{x})(c+1)(f+\bar{x})}{4(a+n/2)(e+\frac{1}{2n}) - (c+1)^2}\right]$$

Ejemplo 4

Los siguientes datos tomados de Rosner (1983), corresponden a una muestra de 52 observaciones que se suponen de una distribución Normal con media desconocida y precisión uno.

-0.28	0.76	1.05	1.29	1.34
1.41	1.41	1.50	1.55	1.60
1.67	1.67	1.74	1.75	1.77
1.85	1.89	1.91	1.97	1.98
2.03	2.14	2.17	2.20	2.23
2.31	2.34	2.35	2.40	2.41
2.50	2.51	2.53	2.63	2.66
2.69	2.77	2.85	2.93	2.96
3.25	3.27	3.27	3.27	3.60
3.60	3.70	4.02	4.12	4.82
5.20	6.74			

El valor 6.74 es declarado como un 'outlier' por Rosner. Escogemos los valores de $a = 1, b = 1, c^2 = 4ae = 4, e = 1, f = 15$. Se escogieron estos valores para que la precisión en las distribuciones marginales de θ y de δ tiendan a cero, es decir, para representar que se tiene un vago conocimiento inicial acerca de los valores de dichos parámetros. El valor de f se escogió así, para controlar la probabilidad de $\delta < 0$ (hacer que sea muy pequeña).

Utilizando el Factor de Bayes para comparar el modelo de no 'outliers' contra el de un 'outlier' cuando θ es desconocida (ecuación (15)) tenemos:

$$Factor = .2244$$

lo cuál nos indica, por ser un valor menor de uno que, en efecto, el valor 6.74 es un 'outlier'.

Barnett y Lewis (1984) recomiendan que cuando se supone una distribución Normal$(0, \theta)$ es recomendable utilizar las estadísticas de Dixon [4].

---

[4] Las estadísticas de la forma

$$y(r, s; p, q) = \frac{X_{(s)} - X_{(r)}}{X_{(q)} - X_{(p)}}$$

llamadas estadísticas de Dixon (Barnett y Lewis, 1984), han sido investigadas por Dixon(1950,1951), Likes(1966) y otros.

En esta sección se verá que los 'outliers' de varias distribuciones pueden ser estudiados por transformaciones a la Exponencial ó a la Normal (Barnett y Lewis, 1984) [5]. Así, podremos hacer uso de los resultados ya obtenidos para dichas distribuciones.

5.4.1) Distribuciones Gumbel, Fréchet y Weibull

Las distribuciones de valores-extremos del primero, segundo y tercer tipo, en otras palabras las distribuciones Gumbel, Fréchet y Weibull, son también conocidas como modelos de observaciones extremas, tales como la máxima velocidad de una partícula, las temperaturas mínimas anuales, las edades más grandes de individuos en una población, los tiempos de vida más cortos de un cierto producto, etc. En el análisis de datos de este tipo es importante detectar posibles 'outliers' en nuestros datos para evitar influencias erróneas en los resultados [6].

La distribución Gumbel depende de un parámetro de localización a y de un parámetros positivo de escala b; las distribuciones Fréchet y Weibull dependen de estos dos parámetros y además de un tercer parámetros positivo de ajuste r. En términos de estos parámetros sus funciones de distribución $P(X < x)$ están dadas en la Tabla 4.

Tabla 4

	Distribución de valores más grandes	Distribución de valores más pequeños
Gumbel	$\exp[-e^{-(x-a)/b}]$, $-\infty < x < \infty$	$1-\exp[-e^{-(a-x)/b}]$, $-\infty < x < \infty$
Fréchet	$\exp\left[-\left(\frac{x-a}{b}\right)^{-r}\right]$, $a < x$	$1-\exp\left[-\left(\frac{a-x}{b}\right)^{-r}\right]$, $x < a$
Weibull	$\exp\left[-\left(\frac{a-x}{b}\right)^{r}\right]$, $x < a$	$1-\exp\left[-\left(\frac{x-a}{b}\right)^{r}\right]$, $a < x$

Cada distribución tiene dos formas, de acuerdo si se refiere a valores-más-grandes o a valores-más-pequeños.

Si X tiene una distribución Gumbel para valores-más-grandes, la transformación

[5] La información de esta sección se obtuvo de Barnett y Lewis (1984), sección 6.4.

[6] Para más información acerca de las distribuciones de valores-extremos se pueden consultar Gnedenko (1943), Gumbel (1958) y Thompson (1969).

$Y = \exp(-x/b)$ tiene una distribución Exponencial con origen 0 y parámetro $\theta = \exp(-a/b)$. Si se conoce el valor de b, podemos transformar los valores observados x_i en $y_i = \exp(x_i/b)$ y usar los resultados de la sección 5.1. Nótese que en esta transformación en particular un 'outlier' superior x_n en la muestra X se convierte a un 'outlier' inferior y_1 en la muestra Y, por tanto los Factores deben usarse de acuerdo a esto.

En la Tabla 5 se muestran las transformaciones correspondientes para las distribuciones Gumbel, Fréchet y Weibull.

En cada caso la distribución de Y tendrá una función de densidad de la forma $\theta \exp(-\theta Y), Y > 0$, y podremos entonces utilizar los resultados de la sección 5.1.

Tabla 5

Distribución de X	Parámetros que se requieren conocer	Transformación a aplicar	Parámetro θ de la distribución de Y	x_n se transforma en	x_1 se transforma en
Gumbel v.m.g.	b conocido	$Y = \exp(-X/b)$	$\exp(a/b)$	$y_{(1)}$	$y_{(n)}$
Gumbel v.m.ch.	b conocido	$Y = \exp(X/b)$	$\exp(-a/b)$	$y_{(n)}$	$y_{(1)}$
Fréchet v.m.g.	a, r conocido	$Y = (X - a)^{-r}$	b^r	$y_{(1)}$	$y_{(n)}$
Fréchet v.m.ch.	a, r conocido	$Y = (a - X)^{-r}$	b^r	$y_{(n)}$	$y_{(1)}$
Weibull v.m.g.	a, r conocido	$Y = (a - X)^r$	b^{-r}	$y_{(1)}$	$y_{(n)}$
Weibull v.m.ch.	a, r conocido	$Y = (X - a)^r$	b^{-r}	$y_{(n)}$	$y_{(1)}$

v.m.g.=valor-mas-grande. v.m.ch.=valor-mas-chico.

5.4.2) Distribución Pareto

Una variable aleatoria X, de la distribución Pareto se caracteriza por dos parámetros, el valor mínimo a, $a > 0$, y el parámetro de ajuste r, $r > 0$. Su función de distribución puede escribirse como

$$\begin{aligned} P(X < x) &= 0, && x \leq a \\ &= 1 - \left(\frac{a}{x}\right)^r, && x \geq a \end{aligned}$$

Si $Y = \ln X$, la función de distribución de Y es

$$\begin{aligned} P(Y < y) &= 0, && y \leq \ln a \\ &= 1 - \exp[-r(y - \ln a)], && y \geq \ln a \end{aligned}$$

y así Y tiene una distribución Exponencial con origen en $\ln a$ y parámetros $\theta = r$.

Si a es conocida, se usa la transformación $Z = \ln(X/a)$. Así los datos transformados como $z_i = \ln[x_i/a]$ tendrán una distribución Exponencial con origen en el 0. Entonces, podemos aplicar los resultados de la sección **5.1**.

5.4.3) Distribución Poisson

Una variable aleatoria X de la distribución Poisson, tiene una función de densidad de la forma

$$f(x \mid \theta) = \frac{\theta^x \exp(-\theta)}{x!}$$

En este caso podemos usar el hecho de que la transformación $\sqrt{(x + 1/4)}$ de una variable aleatoria Poisson X, con media θ se distribuye aproximadamente como una Normal con media $\sqrt{\theta}$ y precisión 4, siempre y cuando θ no sea muy pequeña, digamos mayor que 4 o 5.

De este modo los resultados de la sección **5.3** se pueden usar, ya que por ser la precisión conocida e igual a 4, podemos transformar la muestra a una Normal con precisión igual a 1.

5.4.4) Distribución Binomial

Una variable aleatoria X de la distribución Binomial tiene una función de densidad de la forma

$$f(x \mid \theta, m) = \binom{m}{x} \theta^m (1-\theta)^{m-x}$$

Para una variable aleatoria Binomial con parámetros m, θ la transformación $\operatorname{sen}^{-1}\left[(X/m)^{1/2}\right]$ se distribuye aproximadamente como una Normal con media $\operatorname{sen}^{-1}(\theta)^{1/2}$ y precisión $4m$. Cuando m es conocida y p desconocida podemos aplicar los resultados de la sección **5.3** a la muestra de valores transformados $y_j = \operatorname{sen}^{-1}[(x_j/m)^{1/2}]$.

CONCLUSIONES

En esta tésis se ha expuesto un método Bayesiano para el tratamiento de 'outliers'. A lo largo de la exposición (principalmente **Capítulo 4**) se han hecho varios supuestos. Se ha pedido que las observaciones aberrantes deben ser 'outliers', que se tenga suficiente información inicial acerca de los modelos que genen 'buenas' observaciones y 'outliers' (el modelo generador de 'outliers' debe ser una traslación del modelo que genera 'buenas' observaciones), se deben considerar situaciones en las que sólo se tengan 'outliers' superiores ó situaciones que contengan sólo 'outliers' inferiores y por último, el modelo considerado $f(x \mid \theta)$, debe ser un miembro de la Familia Exponencial con 1 parámetro. Pero también se ha hecho hincapié de que algunas de estas consideraciones se hicieron para facilitar su presentación, pues el método es muy versátil. Además, si nuestro conocimiento inicial acerca del fenómeno nos indica que las condiciones que se piden y las aproximaciones que se hacen son razonables, el método resulta muy útil para el estudio de los 'outliers' y es una buena forma de aprovechar dicho conocimiento inicial.

El método se puede tratar de extender a situaciones que consideren tanto 'outliers' superiores como inferiores, buscando una generalización de las aproximaciones dadas en el **Capítulo 4** para asegurar que las observaciones aberrantes sean 'outliers'. También se puede tratar de extender a miembros de la Familia Exponencial k-parametral. En fin, todavía hay mucho por hacer y por investigar respecto a este método Bayesiano.

Pettit y Smith (1985) hacen notar que en el estudio de los 'outliers' existen muchos problemas. Uno de ellos es un problema conceptual de lo que es un 'outlier'. Tenemos que la idea general de que una observación aberrante es aquella generada por un mecanismo diferente del que genera la mayoría de las observaciones, así sola, sin más elaboración, no engloba la noción de "distanciamiento", que es lo que se entiende por 'outlier'. El método Bayesiano aquí presentado resuelve esta dificultad.

A lo largo de este trabajo, se ha ejemplificado el método Bayesiano usando la distribución Exponencial, la Normal$(0, \theta)$ y la Normal$(\theta, 1)$, pero se puede aplicar a otras distribuciones pertenecientes a la Familia Exponencial con 1-parámetro, como la Binomial, la Poisson y la Gama(n, θ). El hecho de que en la distribución Normal un parámetro sea conocido, no nos restringe a verlo sólo como un ejemplo teórico y nada práctico, pues hay muchas situaciones en las que se pueden conocer. Algunas situaciones donde la media es desconocida pero la precisión es conocida se presenta en problemas de control de calidad, donde la experiencia pasada nos permite dar una valor acertado acerca de la precisión del proceso. Se estudiaron otras distribuciones como la Gumbel, Fréchet, Weibull, Pareto, Binomial y Poisson por medio de transformaciones a la Exponencial y a la Normal.

APENDICES

APENDICE A

A continuación se mencionará el Teorema de Bayes en forma de variables aleatorias y en forma de eventos. (Ver DeGroot (1970, 11-12)

Teorema de Bayes

Sea $X = (X_1, X_2, \ldots, X_m)^T$ y $Y = (Y_1, Y_2, \ldots, Y_n)^T$ dos vectores aleatorios. Sean g_1 y g_2 las funciones de densidad de probabilidad generalizadas (f.d.p.g.) marginales de X y Y respectivamente. Sean $h_1(. \mid y)$ la f.d.p.g. condicional de X dado el valor de $Y = y$ y $h_2(. \mid x)$ la f.d.p.g. condicional de Y dado el valor de $X = x$. Entonces para cualquier punto $x \in R^m$ tal que $g_1(x) > 0$ y cualquier $y \in R^m$

$$h_2(y \mid x) = \frac{h_1(x \mid y) g_2(y)}{\int_{R^n} h_1(x \mid t) g_2(t) d\mu(t)}$$

En forma de eventos el Teorema de Bayes se establece como sigue:

Sean $A_1, A_2, \ldots$ una sucesión de eventos disjuntos para los cuales $\bigcup_{i=1}^{\infty} A_i = S$ donde S es el espacio muestral, es decir, el conjunto de todos los resultados posibles de un experimento, y $\Pr(A_i) > 0$ para $i = 1, 2, \ldots$. Además, sea B otro evento tal que $\Pr(B) > 0$. Entonces:

$$Pr(A_i \mid B) = \frac{Pr(B \mid A_i) Pr(A_i)}{\sum_{j=1}^{\infty} Pr(B \mid A_j) Pr(A_j)} \qquad i = 1, 2, 3, \ldots$$

APENDICE B

Consideremos que en una muestra de tamaño n, $n-1$ observaciones proviene de la distribución

$$f(x \mid \lambda) = \lambda \exp(-\lambda x), \quad x > 0$$

y una observación proviene de la distribución

$$f(x \mid \lambda, \delta) = \delta\lambda \exp(-\delta\lambda x), \quad 0 < \delta < 1$$

Antes de realizar el experimento no se tiene ningún conocimiento inicial de cuál de estas n observaciones es un 'outlier'. Bajo esta consideración la verosimilitud de $(x_1, \ldots, x_n)$ esta dada por

$$f(x_1, \ldots, x_n \mid \lambda, \delta) = \frac{1}{n}\delta\lambda^n \exp\left\{-\sum_{i=1}^{n} \lambda x_i\right\}\left\{\sum_{i=1}^{n} \exp[(1-\delta)\lambda x_i]\right\}$$

Consideremos la variable aleatoria $y_i = \lambda x_i$. Entonces $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ son las estadísticas de orden de

$$f(y_1, \ldots, y_n \mid \delta) = \frac{\alpha}{n} \exp\left\{-\sum_{i=1}^{n} y_i\right\}\left\{\sum_{i=1}^{n} \exp[(1-\delta)y_i]\right\}$$

Sea u_r la probabilidad de que $y_{(r)}$ corresponda a la observación que proviene de $f(y \mid \delta)$.

Entonces u_r se puede valuar como (Kale y Sinha, 1971)

$$\begin{aligned}
u_r &= \sum_{i=1}^{n} \Pr\{y_{(r)} \equiv y_i \mid y_i \sim f(y \mid \delta)\} \Pr\{y_i \sim f(y \mid \delta)\} \\
&= \Pr\{y_{(r)} \equiv y_n \mid y_n \sim f(y \mid \delta)\} \\
&= \binom{n-1}{r-1} \int_0^{\infty} (1-e^{-y})^{r-1} e^{-(n-r)y} \delta e^{-\delta y} dy \\
&= \delta\Gamma(n)\Gamma(n-r+\delta)/\Gamma(n+\delta)\Gamma(n-r+1)
\end{aligned}$$

Como nos interesa $r = n$, tenemos que la probabilidad de que la observación aberrante sea la más grande en la muestra es igual a

$$\frac{\Gamma(n)\Gamma(\delta+1)}{\Gamma(n+\delta)}$$

APENDICE C

La función de densidad Siegel está definida como:

$$Si(k,g,h) = \frac{u\delta^{(g/2)-1}}{B(\frac{1}{2}k,\frac{1}{2}g)h^{g/2}(1+h^{-1}u)^{(k+g)/2}} \qquad u > 0$$

Donde B denota la función Beta como usualmente se define, relacionada con la función Gama por:

$$B(g,h) = \frac{\Gamma(g)\Gamma(h)}{\Gamma(g+h)} \qquad g > 0,\ h > 0$$

Además, se tiene que

$$\int_{a}^{\infty} Si(k,g,h) = I_{h/(h+a)}(\frac{1}{2}k, \frac{1}{2}g)$$

Donde I es la función Beta incompleta definida como:

$$I_a(g,h) = \int_0^a \frac{u^{g-1}(1-u)^{h-1}}{B(g,h)} du$$

APENDICE D

En este capítulo se mencionarán algunos resultados obtenidos al hacerse un análisis de sensitividad en los ejemplos principales del Capítulo 5.

Los resultados del ejemplo 1 de la sección 5.1 fueron:

$\alpha \backslash \beta$	0	0.1	0.4	0.5	1	1.5	2
0	0.130	0.136	0.156	0.162	0.197	0.236	0.278
1	0.125	0.131	0.150	0.156	0.190	0.227	0.268
4	0.110	0.115	0.132	0.138	0.169	0.203	0.240
5	0.105	0.110	0.127	0.132	0.162	0.196	0.232
15	0.085	0.089	0.102	0.107	0.133	0.162	0.194
10	0.068	0.072	0.083	0.088	0.110	0.135	0.162
20	0.055	0.580	0.068	0.071	0.090	0.112	0.136

Vemos que aunque los valores de α y β estén alejados de los valores que generan la muestra, no afectan significativamente los resultados, es decir, siguen apoyando la hipótesis de que el valor 0.459 es un 'outlier'. Con lo que se puede afirmar que para el caso de esta distribución, el método resulta bueno aún cuando nuestro conocimiento inicial no sea muy preciso.

Los resultados del ejemplo 4 de la sección 5.1 fueron:

$b \backslash a$	0.1	0.5	1.0	1.55	10	15	20
1	0.000081	0.000078	0.000075	0.000071	0.000031	0.000019	0.000012
0.5	0.000086	0.000083	0.000079	0.000075	0.000033	0.000021	0.000013
10	0.000092	0.000088	0.000084	0.000080	0.000036	0.000022	0.000014
15	0.000098	0.000095	0.000090	0.000086	0.000039	0.000024	0.000015
20	0.000105	0.000101	0.000097	0.000092	0.000042	0.000026	0.000016

Observamos que al igual que en el ejemplo anterior los valores de a y b iniciales no influyen fuertemente en los resultados, ya que, aunque estén alejados de los valores que generan la muestra, se sigue apoyando la hipótesis de que el valor 92 es un 'outlier'.

Los valores de e y de f fueron 100 y 1, respectivamente. Se escogieron dichos valores para mantener que be^{-1} y $b/[e(a+1)]^{-1}$ sean pequeños, los cuáles se mantienen entre .01 a .2, y .0009 a .18, respectivamente, con los cambios de a y b.

Los resultados del ejemplo 1 de la sección 5.2 fueron:

$\alpha \backslash \beta$	8	4	2	1	.25
1	1.347	0.943	0.676	0.526	0.410
4	1.228	0.825	0.569	0.431	0.326
8	1.086	0.690	0.452	0.329	0.240
16	0.849	0.483	0.286	0.193	0.130
32	0.519	0.237	0.114	0.066	0.038

El valor del parámetro θ que genera la muestra es 4, por lo que los valores de α y β que están cerca de él, apoyan la hipótesis de que el valor .94761 es un 'outlier'. Los valores que apoyan más un valor pequeño para θ dan como resultado que sea dudoso el apoyo de que el valor .94761 se un 'outlier', e incluso se apoyan la hipótesis de que no lo es. Valores que apoyan un valor grande para θ tienden a apoyar que el valor problema sí es un 'outlier'.

Estos resultados van de acuerdo a la lógica de lo que se esperaría, indicándonos que en este caso si no se tiene un conocimiento inicial adecuado es preferible usar una distribución inicial de referencia, que en el caso del ejemplo nos da un valor de:

$$Factor = .401564$$

apoyando ciertamente que el valor .94761 es un 'outlier'.

Los resultados del ejemplo 1 de la sección 5.3 fueron:

β \ α	5	10	15	20
1	0.74240	0.99630	1.33710	1.79420
10	0.07960	0.96940	11.8000	143.000
50	0.00050	0.91119	1647.00	2x10
100	0.00004	0.88337	1945.70	4x10

El valor de α que está de acuerdo con el valor de θ que generó la muestra es 10.

Cuando α es igual a 10, se apoya la hipótesis de que el valor 14.56 es 'outlier', aunque no claramente. Cuando el valor de α y/o β no están de acuerdo con la distribución que genera la muestra los valores se disparan en ambos sentidos, apoyando una y otra hipótesis contrarias.

Los resultados del ejemplo 3 de la sección 5.3 fueron:

β \ α	5	10	15	20
1	0.40605	0.98130	2.37130	5.73040
10	0.00050	0.85150	1539.00	278,344
50	1x10	0.61730	3.6x10	2.2x10
100	4.9x10	0.52560	5.6x10	6x10

Al igual que en ejemplo anterior, se tiene que, cuando los valores de α y/o β no corresponden a los de la distribución que genera los datos, o se tiene una precisión muy baja, el factor se vuelve una forma no muy útil de probar las hipótesis.

En cuanto a los resultados que se obtienen al variar los parámetros, son resultados lógicos, es decir, el modelo es sensible a indicar lo que se debe de acuerdo al conocimiento incial que establezca. Es recomendable usar una distribución inicial de referencia , ya que para obtener buenos resultados es necesario tener un conocimiento inicial casi "exacto" de la "realidad", debido a lo sensible que es el modelo.

Para el ejemplo 1 de la sección 5.3 usando una distribución inicial de referencia se tiene:

$$Factor = .5532,$$

valor que nos apoya la hipótesis de que el valor 14.56 es 'outlier'.

Para el ejemplo 3 de la sección 5.3 usando una distribución inicial de referencia se tiene:

$$Factor = .1680,$$

valor que nos apoya la hipótesis de que los valores en prueba son ambos 'outliers'.

Para terminar recordaremos que, como ya se citó en dicha sección, Barnett y Lewis (1984) recomiendan que cuando se suponga una distribución Normal es preferible utilizar las estadísticas de Dixon.

BIBLIOGRAFIA

Aitchison, J. y Dunsmore, I.R. (1975). *Statistical Prediction Analysis.* Cambridge, Cambridge University Press.

Balasooriya, V. (1989). Detection of Outliers in the Exponential Distribution Based on Prediction. *Commun. Stattis. Theroy Meth..* **18**, 711-720.

Barnett, V. (1983). Marginal outliers in the bivariate Normal distribution. *Bull Int. Stattis. Inst.* **50**, 579-583.

Barnett, V. y Lewis T. (1978). *Outliers in Statistical Data.* 1a. ed. Chichester, Wiley.

Barnett, V. y Lewis T. (1984). *Outliers in Statistical Data.* 2a. ed. Chichester, Wiley.

Beckman y Cook (1983). 'Outlier...s'. *Technometrics.* **25**, 119-163.

Bernardo, J. M. (1979). Expected utility as expected information. *Ann. Statist..* **7**, 686-690.

Bernoulli, D. (1777). Dijudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inductio inde formanda. *Acta Academiai Scientiorium Pertropolitane.* **1**, 3-33.
Traducción a Inglés por C.G. Allen (1961), *Biometrika*, **48**, 3-13.

Bickel, P. J. (1965). On Some Robust Estimates of Locations. *Annals of Mathematical Statistics*. **36**, 347-858.

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *J.R.Statist.Soc.* **A,143**, 383-430.

Box, G. E. P. y Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Massachusetts, Addison-Wesley.

Brumat, C. (1977). el New Develpments in the Applications of Bayesian Methods. (eds. Ahmet, Aykac y Carlo Brumat). Amsterdam, North-Holland.

Chauvenet, W. (1863). Method of least squares. Apéndice del *Manual of Spherical and Practical Astronomy*. Vol. 2. Philadelphia, Lippincott.

Cox, D.R. y Hinkley, D.V. (1974). *Theoretical Statistics*. Londres, Champman and Hall.

De Finetti, (1973). *The Bayesian approach to the rejection of outliers.*

De Groot, M. H. (1970). *Optimal Statistical Decisions*. New York, McGraw-Hill.

De Groot, M. H. (1975). *Probability Statistics*. Massachusetts, Addison-Wesley.

Dixon, W. J. (1950). Analysis of Extreme Values. *Annnals of Mathematical Statistics*, **21**, 488-506.

Dixon, W. J. (1951). Ratios Involving Extreme Values. *Annals of Mathematical Statistics*, **22**, 68-78.

Edgeworth, F.Y. (1883). The method of least squares. *Philosophical Magazine*, **16**, 360-375.

Fisher, R.A. (1929). Test of significance in harmonic analysis. *Proc.R.Soc.* **A,125**, 54-59.

Geisser, S. (1980). Discussion on Sampling and Baye's inference in scientific modelling and robustness. *J.R. Statist.Soc.* **A, 143**, 383-430.

Glaisher (1873). On the rejection of discordant observations. *Monthly Notices Roy. Astr. Soc.*, **33**, 391-402.

Glaisher (1874). Nota sobre una conferencia de Mr. Stone. 'On the rejection of discordant observations'. *Montlhy Notices Roy. Astr. Soc.*, **34**, 251.

Gnedenko, B. V. (1943). Sur la distribution limite de terme maximum d'una série aléatoire. *Ann. Math.* **44**, 423-456.

Goodwin, H. M. (1913). *Elements of the Precision of Measurements and Graphical Methods.* New York, McGraw-Hill.

Gould, B.A.Jr (1855). On Peirce's criterion for the rejection of doubtful observations, with tables for facilitating its application. *Astr. J.* **4**, 81-87.

Grubbs, F.E. (1969). Procedures for detecting outlying observations in samples. *Technometrics.* **11**, 1-21.

Gumbel, E. J. (1985). *Statistics of* extremes. New York, Columbia University Press.

Hogg, R. V. (1965). *Introduction to Mathematical Statistics*, 3a. Ed. McMillan.

Kale y Sinha, S.K. (1971). Estimation of expected life in the presence of an outlier observation. *Technometrics.* **13**, 755-759.

Kimber, A.C. (1982). Test for many outliers in an exponential sample. *Appl. Statist.*, **31**, 263-271.

Kruskal, W. H. (1960). Some remarks on wild observations. *Technometrics*, **2**, 1-3.

Lehman, E. L. (1959). *Testing Statistical Hypotheses.* New York, Wiley.

Likes, J. (1966). Distribution of Dixon's Statistic in the Case of an Exponential Population. *Metrika.* **23**, 27-30.

Lindgren, B. W. (1968). *Statistical Theory.*

Neave, H. R. (1978). *Statistics Tables*. Londres, Allen and Unwin.

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. Ammer. *J. Math.* 8, 343-366.

Peirce, B. (1852). Criterion for the rejection of doubtful observations. *Astr. J.* 2, 161-163.

Pettit, L.I. (1988). Bayes Methods for Outliers in Exponential Samples. *J.R. Statist.Soc.* B, 50, 371-380.

Pettit, L.I. y Smith A.F.M. (1983). Bayesian model comparison in the presence of outliers. *Bull. Statist. Inst.*,50, 292-309.

Pettit, L.I. y Smith A.F.M. (1985). Outliers and influential observations in linear models. En *Bayesian Statistics 2* (eds. J.M.Bernardo, M.H.DeGroot, D.V.Lindley y A.F.M. Smith), pags. 473-494. Amsterdam, North-Holland.

Rosner, R. (1983). Percentage Points for a Generalized ESD Many Outlier Procedure. *Technometrics*. 25, 165-172.

Stone, E.J. (1868). On the rejection of discordant observations. *Monthly Notices Roy. Astr. Soc.*, 28, 165-168.

Thompson, W. A. (1969). *Applied Probability*. New York, Holt, Rinehart and Winston.

Winkler, R. (1972). *Introduction to Bayesian Inference and Decision*. New York, Holt, Rinehart and Winston.

Wright, T. W. (1884). *A Treatise on the Adjustment of Observations by the Method of Least Squares*. New York, Van Nostrand.