

00384

2

24

**INFERENCIA BAYESIANA SOBRE COCIENTES DE COMBINACIONES LINEALES
EN UN MODELO DE REGRESION LINEAL MULTIPLE.**

por

Manuel Mendoza Ramírez

Tesis que se presenta para optar al grado de
Doctor en Ciencias (Matemáticas).

U N A M
1998

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

P R E F A C I O

La investigación que se presenta en esta tesis se originó, como es común en Estadística, en la necesidad de resolver un problema práctico, específico y concreto. En este caso, la revisión documental previa a una aplicación rutinaria de los llamados modelos de *Bioensayo* reveló una serie de peculiaridades en algunos de los procedimientos frecuentistas que se utilizan para obtener inferencias sobre los *parámetros del tipo cociente* que aparecen en estos modelos. Una natural curiosidad motivó la exploración de tratamientos estadísticos alternativos y la posterior identificación de una estructura general que permite incluir a estos modelos dentro de una familia muy amplia, le imprimió un interés adicional al estudio de los posibles procedimientos de análisis estadístico. El producto final es una colección de resultados aplicables en una extensa variedad de problemas que rebasan por mucho el ámbito de los bioensayos y cuyo común denominador es la consideración de parámetros del tipo *cociente*.

Probablemente algún lector perspicaz intuya, después de leer este trabajo, la simpatía del autor por el *enfoque bayesiano*

de la Estadística. En efecto, es difícil pasar por alto los argumentos, especialmente al nivel de fundamentos, que le confieren al procedimiento bayesiano un carácter general, normativo y libre de inconsistencias. De cualquier manera, en este trabajo y adoptando una actitud pragmática, se ha intentado presentar los resultados concretos en la forma más llana y directa posible remitiendo, siempre que ha sido posible, la discusión al contexto particular del problema que se trata y sin recurrir al discurso filosófico.

La tesis está organizada en siete capítulos. En el capítulo 1, se plantea explícitamente el problema de obtener inferencias sobre un cociente de combinaciones lineales de los coeficientes de un modelo de regresión lineal múltiple. El capítulo 2 presenta una revisión del tratamiento frecuentista que recibe este tipo de problema, particularmente en lo que se refiere a la producción de las llamadas regiones de confianza a través del Teorema de Fieller. El capítulo 3 inicia la sección de aportaciones originales del trabajo. En ese capítulo se propone una familia de distribuciones iniciales que, al aplicar el enfoque bayesiano, conducen a distribuciones finales para el parámetro de interés con una serie de interesantes propiedades que en particular evitan las dificultades asociadas a las inferencias frecuentistas.

En el capítulo 4 se prueba que las distribuciones mínimo informativas conocidas como distribuciones de referencia pertenecen a la familia propuesta en el capítulo anterior y para este caso se obtienen algunos resultados más específicos. Colateralmente, en este mismo capítulo 4 se establece un resultado que determina condiciones bajo las cuales la Normalidad asintótica a posteriori de un parámetro se preserva bajo transformaciones. Este resultado es de naturaleza general pero tiene una particular utilidad en la determinación de las distribuciones de referencia.

En los capítulos 5 y 6 se describe la aplicación de los resultados anteriores en los conocidos modelos de cociente de pendientes y líneas paralelas respectivamente. En ambos casos estas aplicaciones se ilustran con ejemplos numéricos. Finalmente, en el capítulo 7 se incluyen algunos comentarios adicionales que se refieren a algunos tópicos relacionados así como a posibles áreas de investigación futura.

El material que se expone en este trabajo fue desarrollado en su totalidad en la Universidad de Valencia, España donde el autor ha tenido el privilegio de disfrutar de la hospitalidad del Departamento de Estadística e Investigación Operativa durante el periodo Octubre 1985 - Junio de 1987. Algunas secciones han sido presentadas, en versiones preliminares, en algunas reuniones científicas. Particularmente se pueden mencionar la I Reunión de la Sección Española de la Biometric Society celebrada en la ciudad de Granada en Mayo de 1986, el International Symposium on Probability and Bayesian Statistics celebrado en Innsbruck, Austria en Septiembre de 1986, el Third International Valencia Meeting on Bayesian Statistics celebrado en Altea, España en Junio de 1987 y el XX Congreso de la Sociedad Matemática Mexicana celebrado en la ciudad de Xalapa en Noviembre de 1987.

Todo este material fue finalmente organizado, y el manuscrito redactado, en el Departamento de Matemáticas de la Facultad de Ciencias de la Universidad Nacional Autónoma de México durante los meses de Agosto - Noviembre de 1987.

Quisiera terminar este prefacio expresando mi reconocimiento al Depto. de Matemáticas de la Facultad de Ciencias que me brindó su apoyo para trasladarme a la Universidad de Valencia y así dedicarme por completo al desarrollo de esta

investigación. También quiero dejar constancia del estímulo que durante este tiempo ha representado para mí el apoyo del Sistema Nacional de Investigadores de México.

La cálida acogida y el estimulante ambiente que me dispensaron en la Universidad de Valencia es algo que jamás podré olvidar. A todos los miembros del Departamento de Estadística les agradezco esa temporada espléndida. Especialmente quiero mencionar a mi supervisor, el Prof. J.M. Bernardo de quien mucho he aprendido y con quien me considero en deuda.

También deseo manifestar aquí que el procesamiento electrónico del texto no hubiese sido tan rápido y eficiente sin la colaboración de Pamela (la micro-computadora de mis amigos Renata y Raul). Finalmente, por su apoyo constante y comprensión quiero dejar constancia de mi profundo agradecimiento para Lelia.

Manuel Mendoza.

México, Diciembre 1987.

CONTENIDO

Prefacio

Capítulo 1. INTRODUCCION. 1

Capítulo 2. ANALISIS FRECUENTISTA. 7

- 2.1 El Teorema de Fieller. 7
- 2.2 El caso del modelo de regresión y la invarianza ante reparametrizaciones. 10
- 2.3. La relación de las regiones de Fieller con el cociente de verosimilitudes generalizado. 16
- 2.4. Los inconvenientes del procedimiento de Fieller. 19

Capítulo 3. ALTERNATIVA BAYESIANA. 26

- 3.1. La función de verosimilitud. 26
- 3.2. La distribución inicial. 28
- 3.3. La distribución final. 33
- 3.4. Invarianza ante reparametrizaciones. 52
- 3.5 Relación con las regiones de Fieller. 70

Capítulo 4. DISTRIBUCION DE REFERENCIA. 81

- 4.1. El análisis de referencia. 81
- 4.2. Normalidad asintótica bajo transformaciones. 87
- 4.3. Distribución inicial de referencia para el cociente de combinaciones lineales. 100

Capítulo 5. EL MODELO DE COCIENTE DE PENDIENTES. 110

- 5.1. El modelo. 116
- 5.2. La distribución final de referencia para ρ . 121
- 5.3. Ejemplos numéricos. 123

Capítulo 6. EL MODELO DE LINEAS PARALELAS. 128

- 6.1. El modelo. 128
- 6.2. La distribución final de referencia para ψ . 134
- 6.3. La distribución final de referencia para ρ . 137
- 6.4. Ejemplos numéricos. 139

Capítulo 7. CONSIDERACIONES FINALES. 152

Referencias.

CAPITULO I. INTRODUCCION

El problema estadístico de producir inferencias sobre un cociente de medias Normales ha sido investigado por diversos autores y desde diferentes enfoques. En su versión más común, este problema puede plantearse como sigue:

Sean X_1, \dots, X_n y Y_1, \dots, Y_m dos muestras aleatorias independientes entre sí, tales que

$$\begin{aligned} X_t &\sim NC(\mu_1, \sigma^2) & ; t = 1, \dots, n \\ Y_j &\sim NC(\mu_2, \sigma^2) & ; j = 1, \dots, m \end{aligned}$$

con μ_1 , μ_2 y σ parámetros desconocidos. Bajo el supuesto de que la media de la segunda muestra no se anula ($\mu_2 \neq 0$), el problema del cociente de medias consiste en producir inferencias acerca del parámetro

$$\rho \equiv \mu_1 / \mu_2$$

que describe la magnitud relativa de las medias. Entre las primeras referencias a este tema se encuentran los artículos de C.I. Bliss (Bliss 1935a, 1935b) en los que se discute la precisión con que se puede ajustar una curva *probit*. La situación en la que

se requiere un *Análisis Estadístico* de este tipo aparece, de manera natural, en el contexto de *Bioensayo*.

Desde un punto de vista estructural, un bioensayo o ensayo biológico puede describirse como un conjunto de estímulos (ocasionalmente con un solo elemento) aplicado bajo circunstancias específicas (condiciones experimentales) y utilizando varios niveles (dosis), a determinadas unidades biológicas (seres humanos, plantas, cultivos, etc.) con la finalidad de producir un conjunto de respuestas observables. El objetivo general de un bioensayo es producir una descripción apropiada de la relación dosis-respuesta involucrada y uno de los problemas particulares más usuales es el de establecer la *potencia relativa* de dos diferentes estímulos. Si se considera un ensayo directo, el parámetro que describe la potencia relativa coincide con un cociente de medias. Por otra parte, el caso de un ensayo indirecto conduce, cuando se utilizan los modelos más comunes (líneas paralelas, cociente de pendientes), a la estimación de la distancia horizontal entre dos rectas paralelas o el cociente de pendientes de dos rectas con ordenada común. En cualquier caso, es claro que se trata de parámetros del mismo tipo cociente.

En la medida en que los bioensayos implican la producción y recolección de datos numéricos, y de la misma forma en que este proceso ocurre en otros campos de investigación, el estudio de las técnicas estadísticas necesarias para el diseño y análisis de un ensayo biológico ha originado la producción de una abundante literatura estadística especializada (Irwin 1937, Berkson 1944, Knudsen y Curtis 1947, Healy 1949, Bliss 1952, Finney 1952/78 y las referencias ahí citadas, entre otros) cuyo propósito es el desarrollo de los procedimientos estadísticos correspondientes. La relación entre Estadística y Bioensayo es particularmente estrecha y como ejemplo basta la siguiente cita

proveniente de Tsutakawa (1982) : " The development of statistical methods in bioassay has paralleled the development of Statistics in general. Not only are the familiar techniques of regression analysis and maximum likelihood extensively used in bioassay, but a number of statistical problems have originated in bioassay and have stimulated the general development of Statistics."

En cualquier caso, y sin lugar a dudas, la referencia clave más importante sobre el tema de métodos estadísticos en bioensayo es el libro de Finney (1952/78) donde se presentan en forma muy completa los métodos estadísticos Frecuentistas para el tratamiento de ensayos biológicos.

Es conveniente señalar que debido a razones de índole histórica y de coincidencia cronológica, los primeros esfuerzos por obtener una solución al problema del cociente de medias - y por extensión a otros problemas que involucran parámetros de este tipo cociente - fueron desarrollados bajo el enfoque frecuentista de la Estadística y desde esa perspectiva, la obtención de estimaciones puntuales ha sido resuelta sin mayores dificultades aparentes.

En la misma manera en que, de acuerdo a los más comunes criterios frecuentistas, la media muestral se considera un estimador aceptable de la media poblacional, el cociente de medias ha sido típicamente estimado por $R = \bar{X} / \bar{Y}$, el correspondiente cociente de medias muestrales. Este procedimiento puede reproducirse si directamente se utilizan los métodos de Momentos o de Máxima Verosimilitud para la derivación de estimadores. Particularmente, las bien conocidas propiedades que, en general, poseen los estimadores de máxima verosimilitud (Kendall y Stuart 1979, capítulo 18) han conducido a una situación en que la estimación puntual de ρ se considera, desde el punto de vista frecuentista, resuelta.

La producción de las llamadas regiones fiduciales o de confianza para ρ es un tema aparte que enfrenta un panorama muy diferente. No solamente ha existido controversia sobre el método más apropiado para obtener estimaciones por intervalos (Fieller 1954, Creasy 1954) sino que el método, basado en el célebre teorema de Fieller, que con el tiempo se ha constituido en la alternativa más popular para resolver este problema adolece de algunos inconvenientes bien conocidos que, aún considerados desde la perspectiva frecuentista, resultan inaceptables en un procedimiento inferencial razonable. Más aún, puesto que el teorema de Fieller, al igual que el método de máxima verosimilitud, se aplica no solamente al caso del cociente de medias sino a toda una clase de problemas que involucran parámetros de este tipo cociente, estas dificultades o inconvenientes resultan de una trascendencia, en la práctica, mayor.

No es el propósito de esta tesis abundar en los argumentos metodológicos y conceptuales que hacen del enfoque Bayesiano una alternativa de consideración inevitable en el proceso general de la Inferencia Estadística. Para obtener una idea bastante completa de este tipo de discusión se pueden consultar los trabajos de Lindley (1971), DeFinetti (1974) y Savage (1977) entre otros. Aquí, desde un punto de vista operativo, basta mencionar que según el paradigma bayesiano, toda la información disponible acerca del parámetro de interés se describe a través de la correspondiente distribución final. En virtud de que el concepto de confianza no existe bajo el enfoque bayesiano, el teorema de Fieller no tiene aplicación alguna y por supuesto, sus inconvenientes no aparecen. En sustitución de las regiones de confianza se pueden considerar las regiones de máxima densidad que están bien definidas para toda distribución propia.

Siendo ese el panorama general, puede mencionarse que concretamente el problema del cociente de medias ha sido considerado desde la perspectiva bayesiana, entre otros, por Kappenman, Geisser y Antle (1970) y Bernardo (1977). Más recientemente, Darby (1980) y Mendoza (1987) han reportado análisis bayesianos para la potencia relativa en los bioensayos de líneas paralelas y de cociente de pendientes respectivamente.

Una inspección de la literatura revela que los resultados bayesianos disponibles para el análisis estadístico de parámetros del tipo cociente están siempre relacionados con problemas muy particulares y típicamente producen conclusiones analíticas solo para el caso en que se utiliza alguna forma de distribución inicial no-informativa. En esta tesis el análisis se extiende para cubrir una situación mucho más general; se considera el problema de obtener inferencias sobre el cociente de combinaciones lineales de los coeficientes de un modelo de regresión lineal múltiple que se puede plantear como sigue:

Sea $\underline{Y} \in \mathbb{R}^N$ un vector aleatorio con distribución Normal multivariada, con vector de medias $X\underline{\theta}$ y matriz de covarianzas $\sigma^2 I$ donde X es una matriz ($N \times k$) de rango completo, $\underline{\theta} \in \mathbb{R}^k$ ($k < N$) es un vector de coeficientes desconocidos y σ es también un parámetro desconocido. Esto es,

$$\underline{Y} \sim NC(X\underline{\theta}, \sigma^2 I) \quad (1.1.1)$$

Bajo estas condiciones el parámetro de interés, sobre el que se requiere obtener inferencias, está definido por

$$\rho \equiv \lambda_1^t \underline{\theta} / \lambda_2^t \underline{\theta} \quad (1.1.2)$$

en donde λ_1 y λ_2 son dos vectores fijos y linealmente

independientes en \mathbb{R}^k tales que $\lambda_2^1 \theta \neq 0$.

Este problema incluye claramente el del cociente de medias e incluye también, entre muchos otros, los problemas de este tipo mencionados anteriormente. Adicionalmente, en este trabajo se propone una familia de distribuciones iniciales que no solo conducen a distribuciones finales con una serie de interesantes propiedades analíticas sino que además, permiten que el científico exprese su conocimiento subjetivo inicial sobre el parámetro de interés θ , en su caso, recurra a una distribución no-informativa.

Antes de concluir esta introducción es conveniente insistir en que si bien se ha recurrido al contexto de bioensayo para ilustrar la importancia que pueden tener los resultados que se presentan en esta tesis, esto se debe a que originalmente el proyecto de investigación apareció de un problema concreto de esta clase pero debe ser claro que su aplicación no se limita en absoluto a este ámbito. Los elementos básicos de un bioensayo pueden identificarse fácilmente con los que aparecen en problemas de otras áreas de investigación (control de calidad, econometría, etc.). Por otra parte existen otros modelos, no necesariamente relacionados con aplicaciones biológicas que pueden ser analizados utilizando los resultados que aquí se presentan. En cualquier forma, en los siguientes capítulos se discutirán con detalle los aspectos fundamentales del procedimiento propuesto y los resultados que se obtienen se ilustrarán con una variedad de ejemplos.

CAPITULO 2. ANALISIS FRECUENTISTA

2.1 El Teorema de Fieller.

Bajo el enfoque frecuentista y de la misma manera en que la estimación puntual de un parámetro como el que se considera en este trabajo, es básicamente resuelta mediante el empleo del método de máxima verosimilitud, el problema de producir las llamadas regiones de confianza para el cociente de medias Normales - y en general para otros parámetros de este tipo cociente - comúnmente se enfrenta mediante el método basado en el Teorema de Fieller (Fieller 1940, 1944 y 1954). El empleo de este método es tan intensivo actualmente que basta una revisión superficial de la literatura para percibir su predominancia. Así, por ejemplo, en un reciente volumen de *Biometrics* (42) se incluyen al menos tres artículos (Srivastava 1986, Williams 1986 y Buonaccorsi 1986) que hacen referencia directa a esta técnica. Siendo esta la situación, es conveniente presentar este método frecuentista y discutir sus principales características. En su versión más simple, el Teorema de Fieller puede enunciarse como sigue:

Teorema 1 (Fieller) Sean a y β dos parámetros tales que $\beta \neq 0$ y sea $p = a / \beta$. Sean a y b respectivamente estimadores insesgados de a y β con distribución Normal y tales que sus

correspondientes varianzas y covarianza están dadas por $\sigma^2 v_{11}$, $\sigma^2 v_{22}$ y $\sigma^2 v_{12}$ con v_{11} , v_{22} y v_{12} valores fijos y conocidos. Sea además, s^2 un estimador insesgado de σ^2 tal que para algún entero q , la magnitud qs^2/σ^2 tiene una distribución $\chi^2_{(q)}$ independiente de a y b . Entonces, una región de confianza con un nivel $(1-p)$ para ρ está dada por

$$R^p = \{ \rho \in \mathbb{R} \mid (a - \rho b)^2 / (s^2 (v_{11} - 2\rho v_{12} + \rho^2 v_{22})) \leq t^2 \}$$

donde t es el cuantil de orden $(1-p/2)$ de una distribución t de Student con q grados de libertad. Equivalentemente, R^p coincide con el intervalo $[R_1, R_2]$ donde

$$R_1 = [R - g \frac{v_{12}}{v_{22}} - \frac{ts}{b} A] / (1 - g)$$

$$R_2 = [R - g \frac{v_{12}}{v_{22}} + \frac{ts}{b} A] / (1 - g)$$

con

$$A = \left\{ v_{11} - 2Rv_{12} + R^2 v_{22} - g \left(v_{11} - \frac{v_{12}^2}{v_{22}} \right) \right\}^{1/2}$$

$$R = a / b$$

$$g = (t^2 s^2 v_{22}) / b^2$$

y t el cuantil de orden $(1-p/2)$ de una distribución t de Student con q grados de libertad.

La demostración del teorema es elemental (cf. Fieller 1954) y se sigue de considerar la distribución de la variable pivotal

$$h(\rho) = (a - \rho b)^2 / [s^2(\rho^2 v_{22} - 2\rho v_{12} + v_{11})] \quad \square$$

Observaciones

- i) R es una estimación puntual natural para ρ .
- ii) Si $v_{12} = 0$, se obtiene una versión simplificada del resultado que aparece en trabajos previos a Fieller, Bliss (1935) por ejemplo.
- iii) Si $g > 1$, la región no es un intervalo sino el complemento de un intervalo.
- iv) Bajo condiciones generales g puede interpretarse como una estadística de prueba para contrastar la hipótesis $\beta = 0$. La región de aceptación con un nivel de significancia p, estaría determinada por la relación $g > 1$.

Para el caso del cociente de medias Normales, tal como se planteó en la introducción, se tiene

$$R = \bar{X} / \bar{Y}$$

$$v_{11} = 1 / n$$

$$v_{22} = 1 / m$$

$$v_{12} = 0$$

$$g = (t^2 s^2) / (m \bar{Y}^2)$$

$$s^2 = \left(\sum_{i=1}^m CX_i - \bar{X} \right)^2 + \sum_{j=1}^m (CY_j - \bar{Y})^2 \quad / \quad (m+n-2)$$

con t el cuantil de orden $(1-p/2)$ de una distribución t de Student con $(m+n-2)$ grados de libertad. Así, los límites de confianza para $\rho = \mu_1 / \mu_2$ están dados por

$$R_1 = \left(\bar{X} / \bar{Y} - Ct_s / \bar{Y} \right) A \quad / \quad (1-g)$$

$$R_2 = \left(\bar{X} / \bar{Y} + Ct_s / \bar{Y} \right) A \quad / \quad (1-g)$$

en donde

$$A = \left\{ (1/n) + (C\bar{X}/\bar{Y})^2 (1/m) - g (1/n) \right\}^{1/2}$$

Es inmediato además, comprobar que la región de aceptación para la hipótesis $\mu_2 = 0$, cuando se utiliza el método de cociente de verosimilitudes generalizado (Kendall y Stuart 1979, capítulo 24) con un nivel de significancia p , está dada por

$$\mathcal{C} = \left\{ X_{(m)}, Y_{(m)} \mid |\bar{Y}| / (Cs/m) < t \right\}$$

o, equivalentemente,

$$\mathcal{C} = \left\{ X_{(m)}, Y_{(m)} \mid g > 1 \right\}.$$

2.2 El caso del modelo de regresión y la invarianza ante reparametrizaciones.

La producción de regiones de confianza para el cociente de combinaciones lineales de los coeficientes de un modelo de regresión lineal múltiple queda resuelta con la siguiente versión del teorema de Fieller:

Teorema 1. Considerese el modelo $\underline{Y} \sim N(\underline{X}\underline{\theta}, \sigma^2 \underline{I})$ con \underline{X} una matriz $(N \times k)$ fija de rango completo $(k < N)$, $\underline{\theta}$ un vector de coeficientes desconocidos y $\sigma > 0$ también desconocido. Una región de confianza, de nivel $(1-\alpha)$, para el parámetro de interés

$$\rho = \frac{\lambda_1^t \underline{\theta}}{\lambda_2^t \underline{\theta}}$$

donde λ_1 y λ_2 son dos vectores fijos y linealmente independientes en \mathbb{R}^k tales que $\lambda_2^t \underline{\theta} \neq 0$, está definida por el intervalo $[R_1, R_2]$ en donde

$$R_1 = \left[R - g \frac{\lambda_1^t (\underline{X}^t \underline{X})^{-1} \lambda_1}{\lambda_2^t (\underline{X}^t \underline{X})^{-1} \lambda_2} - c \left(ts / \lambda_2^t \hat{\underline{\theta}} \right) AC(\hat{\underline{\theta}}) \right] / (1 - g)$$

$$R_2 = \left[R - g \frac{\lambda_1^t (\underline{X}^t \underline{X})^{-1} \lambda_2}{\lambda_2^t (\underline{X}^t \underline{X})^{-1} \lambda_2} + c \left(ts / \lambda_2^t \hat{\underline{\theta}} \right) AC(\hat{\underline{\theta}}) \right] / (1 - g)$$

con

$$AC(\hat{\underline{\theta}}) = \left\{ \left[\lambda_1^t (\underline{X}^t \underline{X})^{-1} \lambda_1 \right] - 2R \left[\lambda_1^t (\underline{X}^t \underline{X})^{-1} \lambda_2 \right] + R^2 \left[\lambda_2^t (\underline{X}^t \underline{X})^{-1} \lambda_2 \right] - g \left[\lambda_1^t (\underline{X}^t \underline{X})^{-1} \lambda_1 \right] - \frac{\left[\lambda_1^t (\underline{X}^t \underline{X})^{-1} \lambda_2 \right]^2}{\left[\lambda_2^t (\underline{X}^t \underline{X})^{-1} \lambda_2 \right]} \right\}^{1/2}$$

y además,

$$R = \frac{\lambda_1^t \hat{\underline{\theta}}}{\lambda_2^t \hat{\underline{\theta}}}$$

$$g = t^2 s^2 \left[\lambda_2^t (\underline{X}^t \underline{X})^{-1} \lambda_2 \right] / \left(\lambda_2^t \hat{\underline{\theta}} \right)^2$$

$$s^2 = (\underline{CY} - \underline{X\hat{\theta}})'(\underline{CY} - \underline{X\hat{\theta}}) / (N - k)$$

Demostración. Basta recordar el teorema de Fieller con

$$a = \underline{\lambda_1^t \hat{\theta}}$$

$$b = \underline{\lambda_2^t \hat{\theta}}$$

$$v_{11} = [\underline{\lambda_1^t (X^t X)^{-1} \lambda_1}]$$

$$v_{12} = [\underline{\lambda_1^t (X^t X)^{-1} \lambda_2}]$$

$$v_{22} = [\underline{\lambda_2^t (X^t X)^{-1} \lambda_2}]$$

$$s^2 = (\underline{CY} - \underline{X\hat{\theta}})'(\underline{CY} - \underline{X\hat{\theta}}) / (N - k)$$

y el resultado se sigue. \square

Esta aplicación del Teorema de Fieller es sencilla, bien conocida y en la literatura puede encontrarse reportada en Zerbe (1978), por ejemplo. Otra forma de reproducir el mismo resultado es mediante el empleo de reparametrizaciones que se ilustra aquí para una posterior comparación con los resultados del siguiente capítulo.

Sea L una matriz ($k \times k$) de rango completo tal que su transpuesta está dada por $L^t = (\underline{l_1}, \dots, \underline{l_k})$ con $\underline{l_1} = \underline{\lambda_1}$ y $\underline{l_2} = \underline{\lambda_2}$ y para la cual, en consecuencia, $\underline{\beta} = L\hat{\theta}$ satisface las relaciones

$$\underline{\beta_1} = \underline{\lambda_1^t \hat{\theta}} \quad \text{y} \quad \underline{\beta_2} = \underline{\lambda_2^t \hat{\theta}}$$

El empleo de una matriz L con estas características es usual en el análisis de modelos lineales (Cf. Scheffé 1959,

capítulo 1). En cualquier caso, es posible reparametrizar el modelo (1.1.1) de manera que

$$\begin{aligned} X\theta &= X(L^{-1}L)\theta \\ &= (XL^{-1})(L\theta) \\ &= Z\beta \end{aligned}$$

con $Z = XL^{-1}$ y, por supuesto, $\beta = L\theta$. Esto es, ahora

$$Y \sim NC(Z\beta, \sigma^2 I) \quad (2.2.1)$$

y el parámetro de interés es simplemente

$$\rho = \beta_1 / \beta_2 \quad (2.2.2)$$

Puesto que los estimadores usuales de β y σ^2 ,

$$\begin{aligned} \hat{\beta} &= (Z'Z)^{-1}Z'Y \\ s_L^2 &= (Y - Z\hat{\beta})'(Y - Z\hat{\beta}) / (N - k) \end{aligned}$$

son independientes y además

$$\begin{aligned} \hat{\beta} &\sim NC(\beta, \sigma^2(Z'Z)^{-1}) \\ (N - k)s_L^2 / \sigma^2 &\sim \chi^2_{(N-k)} \end{aligned}$$

se tiene que si se define $V = (v_{ij}) = (Z'Z)^{-1}$, entonces una región de confianza, de nivel $(1-\alpha)$ para ρ está definida por el intervalo $[R_1, R_2]$ donde

$$R_1 = [\hat{\beta}_1 / \hat{\beta}_2 - g v_{12} / v_{22} - (t_{\alpha/2} / \hat{\beta}_2) AC(\hat{\beta}_1, \hat{\beta}_2)] / (1 - g)$$

$$R_2 = [\hat{\beta}_1 / \hat{\beta}_2 - g v_{12} / v_{22} + (t_{\alpha/2} / \hat{\beta}_2) AC(\hat{\beta}_1, \hat{\beta}_2)] / (1 - g)$$

con

$$AC(\hat{\beta}_1, \hat{\beta}_2) = \{ v_{11} - 2v_{12}\hat{\beta}_1 / \hat{\beta}_2 + v_{22}(\hat{\beta}_1 / \hat{\beta}_2)^2 - g(v_{11} - v_{12}^2 / v_{22}) \}^{1/2}$$

y en donde $g = t_{\alpha/2}^2 v_{12} / \hat{\beta}_2^2$ y t es el cuantil de orden $(1-\alpha/2)$ de una distribución t de Student con $N-k$ grados de libertad.

Para comprobar que los límites en el enunciado del Teorema 1 coinciden con los que se obtienen en el modelo reparametrizado basta observar que

$$\begin{aligned} Z^t Z &= (X L^{-1})^t (X L^{-1}) \\ &= (L^{-1})^t X^t X L^{-1} \end{aligned}$$

de modo que

$$(Z^t Z)^{-1} = L (X^t X)^{-1} L^t$$

y entonces

$$\begin{aligned} Z (Z^t Z)^{-1} Z^t &= (X L^{-1}) (L (X^t X)^{-1} L^t) (X L^{-1})^t \\ &= X (X^t X)^{-1} X^t \end{aligned}$$

En consecuencia,

$$\begin{aligned}
 \hat{\beta} &= (Z^t Z)^{-1} Z^t Y \\
 &= L(X^t X)^{-1} L^t (L^t)^{-1} X^t Y \\
 &= L(X^t X)^{-1} X^t Y \\
 &= L \hat{\beta}
 \end{aligned}$$

Por lo tanto,

$$\hat{\beta}_1 = \lambda_1^t \hat{\beta}$$

$$\hat{\beta}_2 = \lambda_2^t \hat{\beta}$$

Además,

$$\begin{aligned}
 (N-k) s_L^2 &= (Y - Z \hat{\beta})^t (Y - Z \hat{\beta}) \\
 &= Y^t (I - Z(Z^t Z)^{-1} Z^t) Y \\
 &= Y^t (I - X(X^t X)^{-1} X^t) Y \\
 &= (N-k) s^2
 \end{aligned}$$

Finalmente, si e_i es el i -ésimo vector canónico en \mathbb{R}^k , entonces

$$\begin{aligned}
 v_{ij} &= e_i^t (Z^t Z)^{-1} e_j \\
 &= e_i^t [L(X^t X)^{-1} L^t] e_j \\
 &= (e_i^t L) (X^t X)^{-1} (L^t e_j)
 \end{aligned}$$

$$\begin{aligned}
 &= (L^t \underline{\underline{\theta}}_i)^t (X^t X)^{-1} (L^t \underline{\underline{\theta}}_j) \\
 &= \underline{\underline{\lambda}}_i^t (X^t X)^{-1} \underline{\underline{\lambda}}_j
 \end{aligned}$$

de modo que

$$\begin{aligned}
 v_{11} &= \underline{\underline{\lambda}}_1^t (X^t X)^{-1} \underline{\underline{\lambda}}_1 \\
 v_{12} &= \underline{\underline{\lambda}}_1^t (X^t X)^{-1} \underline{\underline{\lambda}}_2 \\
 v_{22} &= \underline{\underline{\lambda}}_2^t (X^t X)^{-1} \underline{\underline{\lambda}}_2.
 \end{aligned}$$

Por lo tanto, no importa como se seleccionen los $k-2$ vectores que completan, con $\underline{\underline{\lambda}}_1$ y $\underline{\underline{\lambda}}_2$, la matriz de rango completo L . En otras palabras, se tiene la siguiente

Proposición 1. *La región de confianza para ρ es invariante ante reparametrizaciones de rango completo.*

2.3 La relación de las regiones de Fieller con el cociente de verosimilitudes generalizado.

Considérese nuevamente el modelo $Y \sim NC(X\underline{\underline{\theta}}, \sigma^2 I)$ con X una matriz $(N \times k)$ fija de rango completo ($k < N$), $\underline{\underline{\theta}}$ un vector de coeficientes desconocidos y $\sigma^2 > 0$ también desconocido. Sean además, ρ_0 un real fijo y $\underline{\underline{\lambda}}_1$ y $\underline{\underline{\lambda}}_2$ dos vectores linealmente independientes, tales que $\underline{\underline{\lambda}}_2^t \underline{\underline{\theta}} \neq 0$. La prueba de la hipótesis

$$H_0 : \left[\frac{\underline{\underline{\lambda}}_1^t \underline{\underline{\theta}}}{\underline{\underline{\lambda}}_2^t \underline{\underline{\theta}}} \right] = \rho_0$$

o, equivalentemente,

$$H_0 : \rho = \rho_0$$

puede efectuarse, utilizando el cociente de verosimilitudes generalizado. Más aún, puede observarse que H_0 es equivalente a un caso particular de la llamada hipótesis lineal general (Searle 1971, sección 3.6). Si $\rho = \rho_0$ se tiene que

$$\left[\frac{\lambda_1^t \theta}{\lambda_2^t \theta} \right] = \rho_0$$

o equivalentemente,

$$\left[\frac{\lambda_1^t \theta}{\lambda_2^t \theta} \right] - \rho_0 \left[\frac{\lambda_2^t \theta}{\lambda_2^t \theta} \right]$$

de manera que H_0 es cierta si y sólo si

$$\left[\frac{\lambda_1^t \theta}{\lambda_2^t \theta} \right] - \rho_0 \left[\frac{\lambda_2^t \theta}{\lambda_2^t \theta} \right] = 0$$

esto es,

$$[\lambda_1 - \rho_0 \lambda_2]^t \theta = 0.$$

Sea $\underline{c} = [\lambda_1 - \rho_0 \lambda_2] \in \mathbb{R}^k$. Puesto que λ_1 y λ_2 son linealmente independientes, se sigue que $\underline{c} \neq \underline{0}$ para todo $\rho_0 \in \mathbb{R}$. Por lo tanto, la prueba de H_0 se puede efectuar probando la hipótesis equivalente

$$H_0^M : \underline{c}^t \theta = 0$$

que es un caso particular de la hipótesis lineal general y tiene asociada, para un nivel de significancia α y utilizando el cociente de verosimilitudes generalizado, la siguiente región de rechazo (Searle 1971, página 112)

$$\bar{\sigma} = \left\{ \underline{Y} \mid \frac{[\underline{e}^t \hat{\underline{\theta}}]^2}{s^2 [\underline{e}^t (\underline{X}^t \underline{X})^{-1} \underline{e}]} > t^2 \right\}$$

donde

$$s^2 = (\underline{Y} - \underline{X} \hat{\underline{\theta}})' (\underline{Y} - \underline{X} \hat{\underline{\theta}}) / (N-k)$$

es el estimador usual de σ^2 y t es el cuantil de orden $(1-\alpha/2)$ de una distribución t de Student con $(N-k)$ grados de libertad. Puesto que

$$\begin{aligned} \underline{e}^t (\underline{X}^t \underline{X})^{-1} \underline{e} &= (\lambda_1 - \rho_0 \lambda_2)' (\underline{X}^t \underline{X})^{-1} (\lambda_1 - \rho_0 \lambda_2) \\ &= (\lambda_1^t - \rho_0 \lambda_2^t)' (\underline{X}^t \underline{X})^{-1} (\lambda_1 - \rho_0 \lambda_2) \\ &= [\lambda_1^t (\underline{X}^t \underline{X})^{-1} - \rho_0 \lambda_2^t (\underline{X}^t \underline{X})^{-1}] (\lambda_1 - \rho_0 \lambda_2) \\ &= [\lambda_1^t (\underline{X}^t \underline{X})^{-1} \lambda_1] - 2\rho_0 [\lambda_1^t (\underline{X}^t \underline{X})^{-1} \lambda_2] \\ &\quad + \rho_0^2 [\lambda_2^t (\underline{X}^t \underline{X})^{-1} \lambda_2] \\ &= \rho_0^2 v_{22} - 2\rho_0 v_{12} + v_{11} \end{aligned}$$

se tiene que

$$\bar{\sigma} = \left\{ \underline{Y} \mid \frac{([\lambda_1^t \hat{\underline{\theta}}] - \rho_0 [\lambda_2^t \hat{\underline{\theta}}])^2}{s^2 [\rho_0^2 v_{22} - 2\rho_0 v_{12} + v_{11}]} > t^2 \right\}$$

o equivalentemente, en términos de la variable pivotal utilizada por Fieller,

$$\bar{\sigma} = \left\{ \underline{Y} \mid h(\rho_0) > t^2 \right\}$$

Esto es, se tiene que para un valor de ρ_0 fijo, y si se denota por Y_N (= \mathbb{R}^N) al espacio muestral, la región de aceptación para la hipótesis H_0^{ρ} está definida por el subconjunto de Y_N

$$\mathcal{A} = \mathcal{A}^{\rho} = \{ Y \in Y_N \mid h(\rho_0) \leq t^2 \}$$

Recordando ahora la definición de \mathbb{R}^{ρ} , la región de confianza para ρ definida en el enunciado del Teorema de Fieller (sección 2.1), es claro que la región de Fieller y la región de aceptación \mathcal{A} guardan una evidente relación. Para ser más precisos, la región de confianza de Fieller puede obtenerse al aplicar el llamado *Método General* para la construcción de regiones de confianza (Rao 1973, página 470) a la familia de eventos

$$A = \{ \mathcal{A} = \mathcal{A}(\rho) \mid \rho \in \mathbb{R} \}$$

Esta relación con las comúnmente aceptadas regiones críticas basadas en el cociente de verosimilitudes generalizado ha contribuido, muy probablemente, a aumentar el impacto de las regiones de confianza de Fieller.

2.4 Los inconvenientes del procedimiento de Fieller.

Las dificultades de las regiones de confianza que se obtienen cuando se emplea el Teorema de Fieller son bien conocidas y de hecho, el autor del procedimiento las reconoce y discute - aunque de manera superficial - en uno de sus artículos básicos (Fieller 1954). Los problemas son básicamente dos y de naturaleza esencialmente distinta. En primer lugar y a un nivel eminentemente

práctico, de acuerdo a la observación (iii) de la sección 2.1, si el cociente g excede el valor de 1 entonces la región de confianza pasa de ser un intervalo a ser su complemento. La utilidad de una afirmación del tipo : " Con una confianza de 95% ρ es menor que R_1 o bien mayor que R_2 " es típicamente escasa. Por otra parte, como se mencionó en la observación (iv) de la misma sección 2.1, usualmente esta situación se puede identificar con aquella en la que la evidencia proporcionada por los datos - juzgada con un criterio frecuentista - no apoya la hipótesis de que el denominador de ρ sea diferente de cero. De esta manera, la dificultad práctica adquiere otra faceta. Para el nivel de significancia - o equivalentemente de confianza - utilizado, la información contenida en los datos sugiere que el parámetro de interés no está bien definido. Desde una perspectiva ingenua, esta dificultad puede salvarse modificando el valor del nivel α . Puesto que

$$g = t^2 s^2_{22} / b^2 \quad (2.4.1)$$

y dado que t^2 es una función monótona decreciente de α ($\alpha \geq 0.5$) es fácil verificar que si α es suficientemente cercana a 0.5 entonces $g < 1$ y la dificultad "desaparece". La falla en este argumento correctivo surge del hecho frecuente de que los valores de α que conducen a un valor de g menor que 1 suelen producir un nivel de confianza tan reducido que a pesar de que la región sea un intervalo, su utilidad es mínima.

El otro inconveniente que tienen las regiones de confianza de Fieller es considerablemente más grave, de carácter conceptual y de hecho, pone en duda la naturaleza misma del concepto de confianza. Como se ha indicado, la región de Fieller es de la forma

$$R^{\alpha} = \{ \rho \in \mathbb{R} \mid h(\rho) \leq t^2 \} \quad (2.4.2)$$

en donde

$$h(\rho) = (a - \rho b)^2 / (s^2(\rho^2 v_{22} - 2\rho v_{12} + v_{11}))$$

y t es el cuantil de orden $(1-\alpha/2)$ de una distribución t de Student con q grados de libertad. Para cada valor fijo de ρ y como función de las variables aleatorias a , b y s^2 , se tiene que $h(\rho)$ sigue una distribución t de Student con los grados de libertad indicados. Es conveniente notar que en forma implícita, el Teorema de Fieller supone que $v_{12}^2 < v_{11}v_{22}$, esto es, que la varianza de $a - \rho b$ es distinta de cero para todo valor de $\rho \in \mathbb{R}$. Esta suposición se puede verificar en cada aplicación particular y típicamente, es correcta. En cualquier caso, la característica más notable de la variable pivotal $h(\rho)$ es que para un conjunto de datos fijo y como función de ρ , es simplemente el cociente de dos polinomios de grado dos. Así, si se definen

$$\begin{aligned} A(\rho) &= (a - \rho b)^2 \\ &= a^2 - 2ab\rho + b^2\rho^2 \\ &= \alpha_2\rho^2 + \alpha_1\rho + \alpha_0 \end{aligned}$$

y

$$\begin{aligned} B(\rho) &= s^2(\rho^2 v_{22} - 2\rho v_{12} + v_{11}) \\ &= \beta_2\rho^2 + \beta_1\rho + \beta_0 \end{aligned}$$

entonces

$$h(\rho) = A(\rho) / B(\rho).$$

Bajo el supuesto de que $v_{12}^2 < v_{11} v_{22}$, el polinomio $B(\rho)$ carece de raíces reales y por tanto la función racional $h(\rho)$, no solo está bien definida para todo $\rho \in \mathbb{R}$ sino que es de clase C^∞ . En particular, es continua y por tanto para todo subconjunto acotado de \mathbb{R} , es a su vez acotada. Finalmente, puesto que cuando $|\rho| \rightarrow \infty$

$$\begin{aligned} \lim h(\rho) &= a_2 / \beta_2 \\ &= b^2 / [s^2 v_{22}] \end{aligned}$$

un real bien definido C que usualmente se puede identificar con la estadística de prueba para la hipótesis $H_0: \beta = 0$, se tiene la siguiente

Proposición 1. *Para un conjunto de datos fijo y como función de ρ , $h(\rho)$ es una función acotada en \mathbb{R} .*

La principal consecuencia de la Proposición 1 es que si $h(\rho)$ es acotada entonces, puesto que t es el cuantil de orden $(1-\alpha/2)$ de una distribución cuyo rango es \mathbb{R} , existe α^* suficientemente cercano a cero, pero estrictamente positivo tal que el valor asociado $(t^*)^2$ es mayor que $h(\rho)$ para todo $\rho \in \mathbb{R}$. Esto es, existe $\alpha^* \in (0,1)$ tal que si t^* es el cuantil de orden $1-\alpha^*/2$ de la distribución t de Student con q grados de libertad, entonces

$$h(\rho) \leq (t^*)^2 \quad \forall \rho \in \mathbb{R}.$$

Por tanto, la correspondiente región de confianza basada en el Teorema de Fieller, con un nivel $(1-\alpha^*/2)$ - estrictamente

menor que 1. puesto que $\alpha^* > 0$ - está dada por

$$\begin{aligned} \mathbb{R}^* &= \{ \rho \in \mathbb{R} \mid h(\rho) \leq (\alpha^*)^2 \} \\ &= \mathbb{R}. \end{aligned}$$

De hecho, el valor de α^* puede determinarse fácilmente si se calculan los valores extremos de $h(\rho)$. Dado que la función es diferenciable en \mathbb{R} , el siguiente procedimiento se aplica

$$h(\rho) = A(\rho) / B(\rho)$$

de donde se tiene que

$$h'(\rho) = [B(\rho)A'(\rho) - A(\rho)B'(\rho)] / [B(\rho)]^2$$

por lo tanto,

$$h'(\rho) = 0 \Leftrightarrow B(\rho)A'(\rho) - A(\rho)B'(\rho) = 0.$$

Ahora bien,

$$A'(\rho) = -2b(a - pb)$$

y entonces

$$B(\rho)A'(\rho) - A(\rho)B'(\rho) = (a - pb)[-2bB(\rho) - (a - pb)B'(\rho)]$$

así que

$$h'(\rho) = 0 \Leftrightarrow (a - pb)[2bB(\rho) + (a - pb)B'(\rho)] = 0$$

esto es si y sólo si ocurre que

$$(a - pb) = 0.$$

o bien

$$2bB(\rho) + (a - pb)B'(\rho) = 0.$$

Por otra parte, $B'(\rho) = 2\beta_2\rho + \beta_1$ de manera que

$$2bB(\rho) + (a - pb)B'(\rho) = (2a\beta_2 + b\beta_1)\rho + (2b\beta_0 + a\beta_1).$$

Más aun, recordando que

$$\beta_2 = s^2v_{22}$$

$$\beta_1 = -2s^2v_{12}$$

$$\beta_0 = s^2v_{11}$$

se tiene que

$$\begin{aligned} 2bB(\rho) + (a - pb)B'(\rho) &= (2s^2av_{22} - 2s^2bv_{12})\rho \\ &+ (2s^2bv_{11} - 2s^2av_{12}) \end{aligned}$$

de manera que $2bB(\rho) + (a - pb)B'(\rho) = 0$ si y sólo si

$$(av_{22} - bv_{12})\rho + (bv_{11} - av_{12}) = 0$$

En resumen, $h'(\rho) = 0$ si y sólo si $\rho = \rho_1$ o bien $\rho = \rho_2$ con

$$\rho_1 = a / b$$

$$\rho_2 = (bv_{11} - av_{12}) / (bv_{12} - av_{22})$$

Puesto que los eventos $b = 0$ y $(bv_{12} - av_{22}) = 0$ ocurren con probabilidad cero, los valores de ρ_1 y ρ_2 están bien definidos con probabilidad uno. Finalmente, si se observa que $h(\rho) \geq 0$, es claro que $\rho = \rho_1$ corresponde a un mínimo ($h(\rho_1) = 0$) y puesto que para $|\rho|$ suficientemente grande es cierto que

$$\frac{h'(\rho)}{h'(-\rho)} > 0$$

se sigue que ρ_2 corresponde a un máximo. Un poco más de álgebra es necesaria para comprobar que el valor máximo de $h(\rho)$ está dado por

$$h(\rho_2) = [b^2 + (av_{22} - bv_{12})^2 / (v_{22}v_{11} - v_{12}^2)] / (s^2v_{22})$$

así que si α se selecciona de forma que $t^2 > h(\rho_2)$, entonces $R^N = R$ como se ha indicado. En resumen, se ha probado la siguiente

Proposición 2. *Para toda aplicación de la técnica de Fieller para la construcción de regiones de confianza existe un valor de α estrictamente positivo tal que la correspondiente región para ρ (con un nivel de confianza estrictamente menor que 1) coincide con toda la recta real.*

Este resultado no solamente podría considerarse como una descalificación del procedimiento de Fieller sino como un inconveniente inherente a las regiones de confianza en general, puesto que como ya se mencionó en la sección 2.3, las regiones de Fieller pueden obtenerse al aplicar un método general de construcción de regiones de confianza.

CAPÍTULO 3. ALTERNATIVA BAYESIANA

3.1 La función de verosimilitud.

Sea, como anteriormente, $\underline{Y} \in \mathbb{R}^N$ un vector aleatorio que sigue una distribución Normal multivariada con vector de medias $X\underline{\theta}$ y matriz de covarianzas $\sigma^2 I$ en donde X es una matriz fija ($N \times k$) de rango completo, $\underline{\theta} \in \mathbb{R}^k$ ($k < N$) es un vector de coeficientes desconocidos y $\sigma > 0$ es también desconocida. Esto es,

$$\underline{Y} \sim N (X\underline{\theta}, \sigma^2 I) \quad (3.1.1)$$

Sea además

$$\rho = \underline{\lambda}_1^t \underline{\theta} / \underline{\lambda}_2^t \underline{\theta} \quad (3.1.2)$$

el parámetro de interés, con $\underline{\lambda}_1$ y $\underline{\lambda}_2$ dos vectores fijos y linealmente independientes en \mathbb{R}^k ($\underline{\lambda}_2^t \underline{\theta} \neq 0$). Recurriendo a las técnicas de reparametrización mencionadas en la sección 2.2 es posible escribir

$$\begin{aligned} X\underline{\theta} &= X (L^{-1}L) \underline{\theta} \\ &= (XL^{-1}) (L \underline{\theta}) \end{aligned}$$

$$= Z \underline{\theta}$$

donde $Z = X L^{-1}$, $\underline{\theta} = L \underline{\beta}$ y L es una matriz ($k \times k$) de rango completo tal que

$$\beta_1 = \lambda_1^t \underline{\theta}$$

$$\beta_2 = \lambda_2^t \underline{\theta}$$

Por tanto, el modelo (3.1.1) puede entonces escribirse como

$$\underline{Y} \sim N (Z \underline{\beta}, \sigma^2 I) \quad (3.1.3)$$

y el parámetro de interés es simplemente

$$\rho = \beta_1 / \beta_2 \quad (3.1.4)$$

La función de densidad del vector de observaciones \underline{Y} está dada por

$$f(\underline{Y} | Z, \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-N/2} \exp \{ -(\underline{Y} - Z \underline{\beta})^t (\underline{Y} - Z \underline{\beta}) / (2\sigma^2) \} \\ \dots (3.1.5)$$

pero una transformación adicional del espacio de parámetros puede aplicarse para expresar esta densidad - y consecuentemente, la correspondiente función de verosimilitud - explícitamente en términos del parámetro de interés ρ . Sea $\underline{\gamma} \in R^k$ un vector tal que

$$\gamma_1 = \rho \\ \gamma_i = \beta_i ; i = 2, \dots, k \quad (3.1.6)$$

Esta es una transformación uno a uno cuya inversa está dada por

$$\begin{aligned} \beta_1 &= \gamma_1 \gamma_2 \\ \beta_i &= \gamma_i \quad ; \quad i = 2, \dots, k \end{aligned} \tag{3.1.7}$$

de manera que β puede expresarse en términos de γ ($\beta = \beta(\gamma)$). Así, la función de verosimilitud de (γ, σ) puede escribirse como

$$L(\gamma, \sigma \mid \underline{Y}) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{(\underline{Y} - Z(\beta(\gamma)))'(\underline{Y} - Z(\beta(\gamma)))}{2\sigma^2}\right\} \tag{3.1.8}$$

que involucra explícitamente a $\rho = \gamma_1$.

Para producir inferencias sobre ρ desde un punto de vista bayesiano, la verosimilitud (3.1.8) debe combinarse con la densidad inicial conjunta $p(\gamma, \sigma)$ que describe el conocimiento sobre γ y σ con que cuenta el científico antes de obtener las observaciones.

3.2 La distribución inicial.

Aplicando el teorema de Bayes, partiendo de una densidad inicial $p(\gamma, \sigma)$ y dado un conjunto específico de observaciones \underline{Y} , la densidad final se obtiene de la relación

$$p(\gamma, \sigma \mid \underline{Y}) \propto p(\gamma, \sigma) L(\gamma, \sigma \mid \underline{Y}) \tag{3.2.1}$$

con $L(\gamma, \sigma \mid \underline{Y})$ la función de verosimilitud de (γ, σ) . En consecuencia, la densidad marginal del parámetro de interés ρ

resulta

$$\begin{aligned} p(\rho | \underline{y}) &= p(\gamma_1 | \underline{y}) \\ &= \int p(\underline{\gamma}, \sigma | \underline{y}) d\gamma_2 \dots d\gamma_k d\sigma \end{aligned} \quad (3.2.2)$$

El objetivo de la densidad inicial $p(\underline{\gamma}, \sigma)$ es describir el estado de conocimiento que se posee sobre los parámetros antes de obtener las observaciones. Es claro por una parte, que distintas representaciones funcionales de esta densidad inicial pueden dar lugar a resultados muy parecidos *a posteriori* y que por otro lado, la dificultad técnica involucrada en el proceso de obtener explícita, analíticamente la densidad marginal (3.2.2) a partir de $p(\underline{\gamma}, \sigma | \underline{y})$ puede ser muy diferente en unos casos y en otros. Así, aparece como un objetivo adicional, secundario pero importante para la elección de la densidad inicial, el empleo de una forma funcional que facilite o simplifique el proceso necesario para obtener la densidad final del parámetro de interés.

Por lo que se refiere al modelo (3.1.8) es conveniente insistir en que está parametrizado por el vector

$$(\underline{\gamma}^1, \sigma) = (\rho, \beta_2, \dots, \beta_k, \sigma)$$

y no solamente por el parámetro de interés ρ . En estas condiciones, β_2, \dots, β_k y σ podrían considerarse como parámetros marginales, de ruido o incidentales que deberían, en algún sentido, ser "eliminados" del análisis. Como ya se ha indicado, la mayor parte de los resultados bayesianos se refieren a casos particulares y han sido derivados utilizando alguna forma de distribución no-informativa conjunta para $(\underline{\gamma}, \sigma)$ o bien, para $(\underline{\beta}, \sigma)$. estos resultados típicamente son complicados desde un punto de vista analítico y requieren de un análisis numérico.

Aquí, el problema de asignar una densidad inicial se enfrentará partiendo del reconocimiento explícito de los distintos papeles que juegan ρ y los parámetros marginales en el modelo. Si la densidad inicial conjunta se escribe como

$$p(\rho, \beta_2, \dots, \beta_k, \sigma) = p(\beta_2, \dots, \beta_k, \sigma | \rho) p(\rho)$$

es claro que la información inicial se puede "descomponer" en dos partes: una, la que se posee sobre el parámetro de interés, descrita por $p(\rho)$, y otra, la que se conoce sobre los parámetros marginales cuando se supone conocido el parámetro de interés y que se describe mediante $p(\beta_2, \dots, \beta_k, \sigma | \rho)$. Respecto a este segundo componente y aún si se cuenta con alguna información subjetiva inicial acerca de los parámetros marginales, puede resultar difícil describir el conocimiento sobre $(\beta_2, \dots, \beta_k, \sigma | \rho)$, especialmente si k es grande. Como alternativa en este trabajo se propone "eliminar" los parámetros marginales utilizando una inicial no-informativa condicional. Para este efecto es útil observar que, para un valor fijo de ρ , el vector de medias $Z(\rho)$ con

$$Z = (z_1, z_2, \dots, z_k) \quad [z_i \in \mathbb{R}^N; i = 1, \dots, k]$$

$$\theta^t = (\beta_1, \beta_2, \dots, \beta_k)$$

puede ser reemplazado por $W \psi$ donde

$$W = (w_1, \dots, w_{k-1}) \quad (N \times (k-1))$$

$$= (\rho z_1 + z_2, z_3, \dots, z_k)$$

$$\psi^t = (\beta_2, \beta_3, \dots, \beta_k)$$

Por lo tanto, condicionado al valor de ρ , el modelo

(3.1.3) puede escribirse como

$$\underline{y} \sim NC W\psi, \sigma^2 I \quad (3.2.3)$$

y una inicial no-informativa condicionada para los parámetros marginales puede obtenerse como el límite de la familia conjugada respectiva (DeGroot 1970, capítulo 9) o utilizando otro método como, por ejemplo, la regla de Jeffreys (Box y Tiao 1973, sección 1.3) o el análisis de referencia (Bernardo 1979). El resultado más importante de esta sección es que para todos estos métodos, la inicial para β_1, \dots, β_k y σ es independiente de ρ y tiene la forma

$$\pi(\beta_1, \dots, \beta_k, \sigma | \rho) \propto \sigma^{-r} \quad (3.2.4)$$

para algún $r > 0$. En consecuencia, en este trabajo se considerará que una densidad inicial conjunta de la forma

$$p(\underline{\chi}, \sigma) \propto p(\rho) \sigma^{-r} \quad (3.2.5)$$

donde $p(\rho)$ es la densidad marginal inicial para ρ que describe el conocimiento que el científico posee sobre el parámetro de interés, provee una representación aproximada razonable para el conocimiento inicial que el científico posee respecto a $(\underline{\chi}, \sigma)$. En la medida en que $p(\rho)$ es una densidad arbitraria, lo mismo que la constante positiva r , es claro que (3.2.5) define una familia de densidades iniciales conjuntas y no una única densidad. Puesto que el vector $(\underline{\chi}^t, \sigma)$ procede de una transformación del vector de parámetros "original" $(\underline{\theta}^t, \sigma)$, resulta de interés establecer la densidad inicial para este vector equivalente a la inicial (3.2.5).

En primer lugar, dado que $\underline{\chi}$ es una transformación de $\underline{\theta}$, si J_χ denota la matriz de derivadas parciales del vector $(\underline{\chi}^t, \sigma)$

respecto al vector $(\underline{\theta}^t, \sigma)$ y recordando que

$$\begin{aligned} \gamma_1 &= \rho \\ &= \beta_1 / \beta_2 \\ \gamma_i &= \beta_i \quad ; i = 1, \dots, k \end{aligned}$$

se tiene que

$$J_1 = \begin{bmatrix} 1/\beta_2 & -\beta_1/\beta_2^2 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & I_{(k-1)} & \dots \\ 0 & 0 & \dots & \dots & \dots & \dots \end{bmatrix}$$

con I_p , la matriz identidad de dimensión p . En consecuencia,

$$|J_1| = 1 / \beta_2 \quad (3.2.6)$$

de donde, aplicando el cambio de variable correspondiente, se tiene que

$$p(\underline{\theta}, \sigma) \propto p_\rho(\beta_1/\beta_2) \sigma^{-r} / |\beta_2| \quad (3.2.7)$$

en donde $p_\rho(\beta_1/\beta_2)$ representa la densidad $p(\rho)$ evaluada en $\rho = \beta_1/\beta_2$. Ahora, si se recuerda que $\underline{\theta} = L\underline{\theta}$, es claro que si J_2 denota la matriz de derivadas parciales de $(\underline{\theta}^t, \sigma)$ respecto al vector $(\underline{\theta}^t, \sigma)$, entonces

$$J_2 = L$$

de manera que $|J_2| = |L|$, independientemente de (θ, σ) . Por tanto, la densidad $p(\chi, \sigma) \propto p(\rho)\sigma^{-r}$ es equivalente a la densidad inicial

$$p(\theta, \sigma) \propto p_{\rho}(\lambda_1^t \theta / \lambda_2^t \theta) \sigma^{-r} / |\lambda_2^t \theta| \quad (3.2.8)$$

en términos de la parametrización "original" (θ, σ) . Claramente, esta densidad no corresponde con las densidades no-informativas que se obtienen vía familias conjugadas o regla de Jeffreys aplicando la técnica correspondiente al vector (θ^t, σ) ; sobre el caso del análisis de referencia se abundará en el siguiente capítulo. Por lo pronto, en la siguiente sección se probará que, sin necesidad de imponer restricciones sobre $p(\rho)$ o sobre la constante positiva r , es posible obtener una expresión analítica para la correspondiente densidad final marginal del parámetro de interés.

3.3 La distribución final.

Utilizando el teorema de Bayes, la densidad conjunta final para (χ, σ) resulta

$$\begin{aligned} p(\chi, \sigma | \underline{Y}) &\propto p(\chi, \sigma) L(\chi, \sigma | \underline{Y}) & (3.3.1) \\ &= [p(\rho) \sigma^{-r}] (2\pi\sigma^2)^{-N/2} \\ &\quad \exp\{-\frac{1}{2\sigma^2}(\underline{Y} - Z\beta(\chi))^t (\underline{Y} - Z\beta(\chi))\} \end{aligned}$$

y por tanto, la densidad marginal final de χ se puede obtener como

$$p(\chi | \underline{Y}) = \int p(\chi, \sigma | \underline{Y}) d\sigma$$

$$\propto \int p(\chi, \sigma) L(\chi, \sigma | \underline{y}) d\sigma$$

$$\propto p(\rho) \int \sigma^{-(N+r)} \exp\{-B/\sigma^2\} d\sigma \quad (3.3.2)$$

con $B = [(\underline{y} - Z(\beta(\chi)))'(\underline{y} - Z(\beta(\chi)))] / 2$. Si en la integral que aparece en (3.3.2) se aplica el cambio de variable $u = \sigma^{-2}$, se obtiene

$$\begin{aligned} \int \sigma^{-(N+r)} \exp\{-B/\sigma^2\} d\sigma &= -\int_0^{\infty} u^{(N+r)/2} \exp\{-Bu\} u^{-3/2} (1/2) du \\ &= (1/2) \int_0^{\infty} u^{(N+r-1)/2} \exp\{-Bu\} du \\ &= (1/2) \int_0^{\infty} u^{A-1} \exp\{-Bu\} du \end{aligned}$$

... (3.3.3)

con $A = (N+r-1)/2$. Claramente, el integrando en (3.3.3) corresponde al núcleo de una densidad Gama con parámetros A y B , de manera que se puede concluir

$$p(\chi | \underline{y}) \propto p(\rho) B^{-A}$$

$$\propto p(\rho) [(\underline{y} - Z(\beta(\chi)))'(\underline{y} - Z(\beta(\chi)))]^{-(N+r-1)/2}$$

... (3.3.4)

Recordando que

$$\begin{aligned} \underline{z}^t &= (\gamma_1, \gamma_2, \dots, \gamma_k) \\ &= (\rho, \gamma_2, \dots, \gamma_k) \\ &= (\rho, \underline{z}_2^t) \end{aligned}$$

con $\underline{\chi}_2^t = (\gamma_2, \dots, \gamma_k)$, se tiene que ahora es necesario calcular

$$\begin{aligned} p(\rho | \underline{Y}) &= p(\gamma_1 | \underline{Y}_2) \\ &= \int p(\underline{\chi} | \underline{Y}) d\underline{\chi}_2 \\ &\propto p(\rho) \int [(Y - Z\underline{\beta}(\underline{\chi}))^t (Y - Z\underline{\beta}(\underline{\chi}))]^{-N/2} d\underline{\chi}_2 \\ &\dots (3.3.5) \end{aligned}$$

La evaluación de la integral en (3.3.5) puede efectuarse observando que es necesario integrar con respecto a β_2, \dots, β_k para cada valor fijo de ρ . En estas condiciones, se puede recurrir a la expresión alternativa (3.2.3) para simplificar el procedimiento. Más específicamente, sean $\underline{z}_1, \underline{z}_2, \dots, \underline{z}_k$ los vectores columna que forman la matriz Z . Esto es,

$$Z = (\underline{z}_1, \underline{z}_2, \dots, \underline{z}_k) \quad (N \times k)$$

y sea además, W la matriz de dimensiones $(N \times (k-1))$ definida en la sección 3.2 cuyas columnas $\underline{w}_1, \underline{w}_2, \dots, \underline{w}_{k-1}$ satisfacen las relaciones

$$\begin{aligned} \underline{w}_1 &= \underline{z}_2 + \rho \underline{z}_1 \\ \underline{w}_t &= \underline{z}_{t+1} \quad ; t = 2, \dots, k-1 \end{aligned}$$

Entonces

$$\begin{aligned} \underline{Y} - Z\underline{\beta}(\underline{\chi}) &= \underline{Y} - \sum_{j=1}^k \beta_j(\underline{\chi}) \underline{z}_j \\ &= \underline{Y} - \{ \beta_1(\underline{\chi}) \underline{z}_1 + \beta_2(\underline{\chi}) \underline{z}_2 + \dots + \beta_k(\underline{\chi}) \underline{z}_k \} \end{aligned}$$

De las ecuaciones (3.1.7) se tiene que

$$\begin{aligned} \beta_1(\chi) &= \gamma_1 \gamma_2 \\ &= \rho \gamma_2 \\ \beta_i(\chi) &= \gamma_i \quad ; \quad i = 2, \dots, k \end{aligned}$$

de manera que

$$\begin{aligned} \underline{Y} - Z(\chi)\chi &= \underline{Y} - \{ \rho \gamma_2 z_1 + \gamma_2 z_2 + \gamma_3 z_3 + \dots + \gamma_k z_k \} \\ &= \underline{Y} - \{ \gamma_2 (z_2 + \rho z_1) + \gamma_3 z_3 + \dots + \gamma_k z_k \} \\ &= \underline{Y} - \{ \gamma_2 w_1 + \gamma_3 w_2 + \dots + \gamma_k w_{k-1} \} \\ &= \underline{Y} - \sum_{j=1}^{k-1} \gamma_{j+1} w_j \\ &= \underline{Y} - W \chi_2 \end{aligned}$$

Por lo tanto, para cada valor fijo de $\rho \in \mathbb{R}$,

$$(\underline{Y} - Z(\chi)\chi)'(\underline{Y} - Z(\chi)\chi) = (\underline{Y} - W\chi_2)'(\underline{Y} - W\chi_2) \quad (3.3.6)$$

con W una matriz $(N \times (k-1))$ fija y de rango completo. Sea ahora

$$\hat{\chi}_2 = (W'W)^{-1} W'\underline{Y}$$

el estimador usual de mínimos cuadrados de χ_2 en el modelo

$$\underline{Y} \sim N(\underline{Y}, \sigma^2 I)$$

Es bien conocido (Searle 1971, página 113) que para todo vector

$$\chi_2 \in \mathbb{R}^{k-1}.$$

$$\begin{aligned} (\underline{Y} - W\chi_2)'(\underline{Y} - W\chi_2) &= (\underline{Y} - W\hat{\chi}_2)'(\underline{Y} - W\hat{\chi}_2) \\ &\quad + (\chi_2 - \hat{\chi}_2)'(W'W)(\chi_2 - \hat{\chi}_2) \end{aligned}$$

... (3.3.7)

de manera que esta suma de cuadrados puede expresarse como sigue

$$(\underline{Y} - W\chi_2)'(\underline{Y} - W\chi_2) = [1 + (\chi_2 - \hat{\chi}_2)'(W'W)(\chi_2 - \hat{\chi}_2)/HC(\rho)] HC(\rho)$$

... (3.3.8)

donde

$$HC(\rho) = (\underline{Y} - W\hat{\chi}_2)'(\underline{Y} - W\hat{\chi}_2)$$

es independiente de χ_2 . La relación (3.3.8) es válida siempre que $HC(\rho) \neq 0$ pero es una consecuencia inmediata de los resultados básicos de modelos lineales que $p[HC(\rho) = 0] = 0$ si $N \geq k$; por lo tanto, esta suposición no es en absoluto restrictiva. De esta forma se tiene que

$$\begin{aligned} p(\rho|\underline{Y}) &= p(\chi_1 | \underline{Y}) \\ &= \int p(\chi_2 | \underline{Y}) d\chi_2 \\ &\propto p(\rho) [HC(\rho)]^{-c/2} \int [1 + (\chi_2 - \hat{\chi}_2)'(W'W)(\chi_2 - \hat{\chi}_2)/HC(\rho)]^{-c/2} d\chi_2 \\ &= p(\rho) [HC(\rho)]^{-c/2} \int [1 + (\chi_2 - \hat{\chi}_2)'T(\chi_2 - \hat{\chi}_2)]^{-c/2} d\chi_2 \end{aligned}$$

... (3.3.9)

con $c = N+r-1$ y $T = (W^t W) / HC\rho$. Resulta inmediato que el integrando en (3.3.9) coincide con el núcleo de una densidad t de Student multivariada con $N+r-k$ grados de libertad, vector de localización \hat{Z}_2 y matriz de precisión T . Por lo tanto, se tiene que (DeGroot 1970, página 61) como función de ρ ,

$$\int (1 + (Z_2 - \hat{Z}_2)^t T (Z_2 - \hat{Z}_2))^{-c/2} dZ_2 \propto |T|^{-1/2} \quad \dots(3.3.10)$$

Ahora bien, $(W^t W)$ es una matriz de dimensión $(k-1) \times (k-1)$ de manera que

$$\begin{aligned} |T| &= |W^t W / HC\rho| \\ &= |W^t W| [HC\rho]^{-(k-1)} \end{aligned}$$

en consecuencia,

$$\begin{aligned} p(\rho | \underline{y}) &\propto p(\rho) [HC\rho]^{-c/2} [HC\rho]^{(k-1)/2} |W^t W|^{-1/2} \\ &= p(\rho) [HC\rho]^{-(N+r-k)/2} |W^t W|^{-1/2} \quad (3.3.11) \end{aligned}$$

Por otra parte,

$$\begin{aligned} W &= (w_1, w_2, \dots, w_{k-1}) \\ &= (z_2 + \rho z_1, z_3, \dots, z_k) \\ &= (w_1, z_2) \end{aligned}$$

con Z_2 la matriz $(N \times (k-2))$ cuyas columnas son z_3, \dots, z_k ($k \geq 3$). En consecuencia,

$$\begin{aligned}
 w^t w &= \begin{bmatrix} w_1^t \\ z_2^t \end{bmatrix} (w_1, z_2) \\
 &= \begin{bmatrix} w_1^t w_1 & \dots & w_1^t z_2 \\ \dots & \dots & \dots \\ z_2^t w_1 & \dots & z_2^t z_2 \end{bmatrix}_{(k-1) \times (k-1)} \quad (3.3.12)
 \end{aligned}$$

en donde

$$\begin{aligned}
 w_1^t w_1 &= (z_2 + \rho z_1)^t (z_2 + \rho z_1) \\
 &= (z_2^t + \rho z_1^t) (z_2 + \rho z_1) \\
 &= \rho^2 (z_1^t z_1) + 2\rho (z_1^t z_2) + (z_2^t z_2) \quad ; (1 \times 1) \\
 w_1^t z_2 &= (z_2 + \rho z_1)^t z_2 \\
 &= \rho (z_1^t z_2) + (z_2^t z_2) \quad ; (1 \times (k-2)) \\
 z_2^t w_1 &= (w_1^t z_2)^t \\
 &= \rho (z_2^t z_1) + (z_2^t z_2) \quad ; ((k-2) \times 1)
 \end{aligned}$$

En cualquier caso se puede comprobar, utilizando los resultados de matrices particionadas (Searle 1982, página 258) que

$$\begin{aligned}
 |w^t w| &= |z_2^t z_2| |w_1^t w_1 - (w_1^t z_2) (z_2^t z_2)^{-1} (z_2^t w_1)| \\
 &= |z_2^t z_2| |w_1^t w_1 - w_1^t z_2 (z_2^t z_2)^{-1} z_2^t w_1| \\
 &= |z_2^t z_2| w_1^t (I - Z_2 (Z_2^t Z_2)^{-1} Z_2^t) w_1 \\
 &= |z_2^t z_2| w_1^t (I - P) w_1 \quad (3.3.13)
 \end{aligned}$$

con $I - P = (I - Z_2(Z_2^t Z_2)^{-1} Z_2^t)$, la matriz simétrica, idempotente y positiva definida asociada a la suma de cuadrados del error en un modelo de la forma

$$\underline{Y} \sim NC(Z_2 \underline{\alpha}, \sigma^2 I) \quad ; \quad \underline{\alpha} \in \mathbb{R}^{(k-2)}$$

Es evidente que, en particular, $\underline{w}_1^t (I - P) \underline{w}_1 \neq 0$ puesto que para cualquier valor de $\rho \in \mathbb{R}$, \underline{w}_1 es una combinación lineal de las dos primeras columnas de Z y entonces, la única forma posible de que esta combinación lineal fuese ajustada con error cero por un modelo lineal en términos de Z_2 es que Z fuese singular. Sin embargo, por hipótesis, Z es de rango completo. En resumen, de (3.3.11) y (3.3.13) se sigue que

$$p(\rho | \underline{Y}) \propto p(\rho) [h(\rho)]^{-(N+r-k)/2} \{\underline{w}_1^t (I - P) \underline{w}_1\}^{-1/2}$$

... (3.3.14)

Un poco más de cálculo permite obtener la expresión alternativa

$$\begin{aligned} \underline{w}_1^t (I - P) \underline{w}_1 &= \rho^2 [\underline{z}_1^t (I - P) \underline{z}_1] + 2\rho [\underline{z}_1^t (I - P) \underline{z}_2] \\ &\quad + [\underline{z}_2^t (I - P) \underline{z}_2] \\ &= Q(\rho) \end{aligned} \tag{3.3.15}$$

con $Q(\rho) = c_2 \rho^2 + c_1 \rho + c_0$ en donde

$$c_2 = \underline{z}_1^t (I - P) \underline{z}_1$$

$$c_1 = 2 \underline{z}_1^t (I - P) \underline{z}_2$$

$$c_0 = \underline{z}_2^t (I - P) \underline{z}_2$$

Las propiedades del polinomio $Q(\rho)$ pueden establecerse en forma inmediata. Exactamente por el mismo razonamiento que se siguió para el caso de \underline{w}_1 , se tiene que

$$c_2 > 0$$

$$c_0 > 0$$

por consiguiente, $Q(\rho)$ siempre es un polinomio de grado exactamente 2. Más aún, el discriminante de la ecuación

$$Q(\rho) = 0$$

está dado por

$$\begin{aligned} c_1^2 - 4c_2c_0 &= \{2 [\underline{z}_1^t(I-P)\underline{z}_2]\}^2 - 4[\underline{z}_1^t(I-P)\underline{z}_1][\underline{z}_2^t(I-P)\underline{z}_2] \\ &= 4\{ [\underline{z}_1^t(I-P)\underline{z}_2]^2 - [\underline{z}_1^t(I-P)\underline{z}_1][\underline{z}_2^t(I-P)\underline{z}_2] \} \end{aligned}$$

Puesto que $(I-P)$ es idempotente, se tiene que si se definen

$$\underline{v}_1 = (I-P)\underline{z}_1$$

$$\underline{v}_2 = (I-P)\underline{z}_2$$

entonces

$$c_1^2 - 4c_2c_0 = 4\{ (\underline{v}_1^t \underline{v}_2)^2 - (\underline{v}_1^t \underline{v}_1)(\underline{v}_2^t \underline{v}_2) \}$$

Por lo tanto, una aplicación inmediata de la desigualdad de Schwarz revela que $c_1^2 - 4c_2c_0 \leq 0$ y que la igualdad se alcanza si y sólo si existe $a \in \mathbb{R}$ tal que

$$\underline{y}_1 = a\underline{y}_2$$

Sin embargo

$$\underline{y}_1 = a\underline{y}_2$$

$$\Leftrightarrow \underline{y}_1 - a\underline{y}_2 = \underline{0}$$

$$\Leftrightarrow (I-P)\underline{z}_1 - (I-P)\underline{z}_2 = \underline{0}$$

$$\Leftrightarrow (I-P)(\underline{z}_1 - a\underline{z}_2) = \underline{0}$$

$$\Leftrightarrow (\underline{z}_1 - a\underline{z}_2)^t (I-P) (\underline{z}_1 - a\underline{z}_2) = 0$$

De manera que nuevamente por las mismas causas por las que $\underline{w}_1^t (I - P) \underline{w}_1 \neq 0$ y puesto que \underline{z}_1 y \underline{z}_2 son linealmente independientes, se tiene que $c_1^2 - 4c_2c_0 < 0$. En conclusión, se ha probado que $Q(\rho)$ es un polinomio de grado 2, sin raíces reales y positivo. En cualquier caso, volviendo al argumento principal se tiene que

$$p(\rho | \underline{y}) \propto p(\rho) [Q(\rho)]^{-1/2} [HC(\rho)]^{-(N+r-k)/2} \quad (3.3.16)$$

en donde $Q(\rho) = c_2\rho^2 + c_1\rho + c_0$ con

$$c_2 = \underline{z}_1^t (I-P) \underline{z}_1$$

$$c_1 = 2 \underline{z}_1^t (I-P) \underline{z}_2$$

$$c_0 = \underline{z}_2^t (I-P) \underline{z}_2$$

y además

$$HC(\rho) = (Y - W\hat{\chi}_2)^t (Y - W\hat{\chi}_2) \quad (3.3.17)$$

Es conveniente observar que en la expresión (3.3.13) para el determinante de la matriz $W^t W$ se ha utilizado implícitamente el hecho de que $k \geq 3$. El caso $k = 2$ conduce a la expresión alternativa

$$\begin{aligned} |W^t W| &= w_1^t w_1 \\ &= \rho^2 (\sum_1^t \sum_1) + 2\rho (\sum_1^t \sum_2) + (\sum_2^t \sum_2) \end{aligned}$$

que tiene asociado un análisis similar pero más simple. En lo que resta, por conveniencia se mantendrá la hipótesis $k \geq 3$; de cualquier forma, el caso $k = 2$ requiere de modificaciones menores y siempre resulta más simple.

Un primer examen puede sugerir que la expresión (3.3.17) para $HC(\rho)$, a diferencia del polinomio $Q(\rho)$, es una función considerablemente complicada de ρ cuya evaluación puede resultar costosa en un problema real. Tal no es el caso. En realidad, como se mostrará en seguida, $HC(\rho)$ es un cociente de dos polinomios de grado dos. Considérese nuevamente el modelo

$$Y \sim NC(W\chi_2, \sigma^2 I)$$

con

$$W = (w_1, w_2, \dots, w_{k-1}) \quad ; \quad (N \times (k-1))$$

$$\chi_2^t = (\gamma_2, \dots, \gamma_k) \in \mathbb{R}^{k-1}$$

y las particiones

$$\begin{aligned}
 W &= (W_1, W_2) \\
 &= (W_1, Z_2) \\
 Z_2^t &= (\gamma_2, Z_2^t) \\
 Z_2^t &= (\gamma_3, \gamma_4, \dots, \gamma_k) \in \mathbb{R}^{k-2}
 \end{aligned}$$

Sea ahora H la hipótesis de que $\gamma_2 = 0$. Esto es

$$H: \underline{e}_1^t Z_2 = 0$$

donde, como es usual, \underline{e}_1 es el primer vector de la base canónica en \mathbb{R}^{k-1} ; $\underline{e}_1^t = (1, 0, \dots, 0)$. Entonces, utilizando la terminología del enfoque frecuentista para la prueba de hipótesis en el modelo lineal general, se tiene que el llamado *Modelo Completo* (MC) está dado por

$$\underline{Y} \sim NC(WZ_2, \sigma^2 I)$$

mientras que el *Modelo Reducido* coincide con

$$\underline{Y} \sim NC(Z_2 Z_2, \sigma^2 I) \quad (3.3.18)$$

y las correspondientes sumas de cuadrados del error satisfacen la ecuación fundamental

$$SCE_{MR} = SCE_{MC} + SCE_H \quad (3.3.19)$$

en donde (Searle 1971, página 113),

$$SCE_H = (\underline{e}_1^t \hat{Z}_2)^t [\underline{e}_1^t (W^t W)^{-1} \underline{e}_1]^{-1} (\underline{e}_1^t \hat{Z}_2)$$

Sea ahora $U = (u_{ij}) \equiv (W^t W)^{-1}$. Utilizando un resultado estándar de álgebra y la representación de $W^t W$ como matriz particionada - expresión (3.3.12) - se tiene que

$$\begin{aligned} e_1^t (W^t W)^{-1} e_1 &= e_1^t U e_1 \\ &= u_{11} \\ &= |Z_2^t Z_2| / |W^t W| \end{aligned} \quad (3.3.20)$$

como además $e_1^t \hat{\gamma}_2 = \hat{\gamma}_2$ es un escalar, se sigue que

$$SCE_M = |W^t W| (\hat{\gamma}_2)^2 / |Z_2^t Z_2|$$

pero de (3.3.13) y (3.3.15) se sabe que

$$|W^t W| = \alpha(\rho) |Z_2^t Z_2|$$

por lo tanto

$$\begin{aligned} SCE_M &= \alpha(\rho) |Z_2^t Z_2| (\hat{\gamma}_2)^2 / |Z_2^t Z_2| \\ &= \alpha(\rho) (\hat{\gamma}_2)^2 \end{aligned} \quad (3.3.21)$$

Más aún, puesto que $\hat{\gamma}_2 = e_1^t \hat{\gamma}_2$, basta con calcular la primera entrada del vector $\hat{\gamma}_2$, donde

$$\hat{\gamma}_2 = (W^t W)^{-1} W^t Y$$

o equivalentemente

$$(W^t W) \hat{\gamma}_2 = W^t Y \quad (3.3.22)$$

Ahora bien, en virtud de que $W = (w_1, Z_2)$ se tiene que

$$\begin{aligned} W^t Y &= \begin{bmatrix} w_1^t \\ Z_2^t \end{bmatrix} Y \\ &= \begin{bmatrix} w_1^t Y \\ Z_2^t Y \end{bmatrix} \end{aligned} \quad (3.3.23)$$

y si se define

$$V_0 = \begin{bmatrix} w_1^t Y & \dots & w_1^t Z_2 \\ Z_2^t Y & \dots & Z_2^t Z_2 \end{bmatrix} \quad (k-1) \times (k-1) \quad (3.3.24)$$

es inmediato comprobar que, utilizando el método de determinantes de Cramer,

$$\begin{aligned} \hat{Z}_2 &= |V_0| / |W^t W| \\ &= |V_0| / [(Z_2^t Z_2) |Q(\rho)|] \end{aligned} \quad (3.3.25)$$

Por otra parte (Morrison 1970, página 68).

$$\begin{aligned} |V_0| &= |Z_2^t Z_2| \left| w_1^t Y - (w_1^t Z_2) (Z_2^t Z_2)^{-1} (Z_2^t Y) \right| \\ &= |Z_2^t Z_2| w_1^t (I - Z_2 (Z_2^t Z_2)^{-1} Z_2^t) Y \\ &= |Z_2^t Z_2| w_1^t (I - P) Y \\ &= |Z_2^t Z_2| (Z_2 + \rho Z_1)^t (I - P) Y \\ &= |Z_2^t Z_2| (Z_2^t + \rho Z_1^t) (I - P) Y \end{aligned}$$

$$\begin{aligned}
 &= [Z_2^t Z_2] \{ (Z_2^t I - P) Y \} + \rho [Z_1^t (I - P) Y \} \\
 &= [Z_2^t Z_2] \{ S_{SV2} + \rho S_{SV1} \} \quad (3.3.26)
 \end{aligned}$$

con $S_{SV2} = Z_2^t (I - P) Y$ y $S_{SV1} = Z_1^t (I - P) Y$. Por lo tanto, de las expresiones (3.3.25) y (3.3.26) se tiene que

$$\begin{aligned}
 \hat{Z}_2 &= [Z_2^t Z_2] \{ S_{SV2} + \rho S_{SV1} \} / [Z_2^t Z_2] (\alpha \rho) \\
 &= (S_{SV2} + \rho S_{SV1}) / (\alpha \rho) \quad (3.3.27)
 \end{aligned}$$

En consecuencia

$$\begin{aligned}
 SCE_M &= (\alpha \rho) (\hat{Z}_2)^2 \\
 &= (\alpha \rho) [S_{SV2} + \rho S_{SV1}]^2 / (\alpha \rho)^2 \\
 &= [S_{SV2} + \rho S_{SV1}]^2 / (\alpha \rho) \quad (3.3.28)
 \end{aligned}$$

Por otra parte

$$SCE_{MR} = (Y - Z_2 \hat{Z}_2)^t (Y - Z_2 \hat{Z}_2) \quad (3.3.29)$$

con $\hat{Z}_2 = (Z_2^t Z_2)^{-1} Z_2^t Y$. De esta manera,

$$\begin{aligned}
 SCE_{MR} &= Y^t [I - Z_2^t (Z_2^t Z_2)^{-1} Z_2^t] Y \\
 &= Y^t (I - P) Y \\
 &= S_Y^2 \quad (3.3.30)
 \end{aligned}$$

Finalmente entonces,

$$\begin{aligned}
 HC(\rho) &= (\underline{Y} - W\hat{Z}_2)'(\underline{Y} - W\hat{Z}_2) \\
 &= SCE_{MC} \\
 &= SCE_{MR} - SCE_N \\
 &= \underline{Y}'(CI - P)\underline{Y} - \{Z_2'(CI - P)\underline{Y} + \rho Z_1'(CI - P)\underline{Y}\}^2 / Q(\rho) \\
 &= S_Y^2 - \{S_{ZY2} + \rho S_{ZY1}\}^2 / Q(\rho) \quad (3.3.31) \\
 &= [Q(\rho)S_Y^2 - \{S_{ZY2} + \rho S_{ZY1}\}^2] / Q(\rho) \quad (3.3.32) \\
 &= J(\rho) / Q(\rho) \quad (3.3.33)
 \end{aligned}$$

con $J(\rho) = Q(\rho)S_Y^2 - \{S_{ZY2} + \rho S_{ZY1}\}^2$. De esta forma, resulta claro que $HC(\rho)$ es un cociente de dos polinomios de grado dos que involucra nuevamente a la función $Q(\rho)$. La expresión para la densidad marginal final de ρ puede reescribirse ahora como

$$p(\rho | \underline{Y}) \propto p(\rho) [Q(\rho)]^{-1/2} [HC(\rho)]^{-(N+r-k)/2} \quad (3.3.34)$$

$$\begin{aligned}
 &= p(\rho) [Q(\rho)]^{-1/2} [S_Y^2 - \{S_{ZY2} + \rho S_{ZY1}\}^2 / Q(\rho)]^{-(N+r-k)/2} \\
 &\quad \dots (3.3.35)
 \end{aligned}$$

$$\begin{aligned}
 &= p(\rho) [Q(\rho)]^{(d-1)/2} [Q(\rho)S_Y^2 - \{S_{ZY2} + \rho S_{ZY1}\}^2]^{-d/2} \\
 &\quad \dots (3.3.36)
 \end{aligned}$$

$$= p(\rho) [Q(\rho)]^{(d-1)/2} [J(\rho)]^{-d/2} \quad (3.3.37)$$

en donde $d = N+r-k$. En cualquier caso, a partir de la expresión $HC(\rho) = (\underline{Y} - W\hat{Z}_2)'(\underline{Y} - W\hat{Z}_2)$ y recordando que $N > k-1$, se sigue que

$p\{H(\rho) = 0\} = 0$ o equivalentemente, $p\{J(\rho) = 0\} = 0$. En consecuencia $p(\rho | \underline{Y})$ está bien definida para toda $\rho \in \mathbb{R}$ con probabilidad 1. Más aun, puesto que $[H(\rho)]^{-1}$ es también un cociente de polinomios de grado dos cuyo denominador no tiene raíces reales, se sigue que $[H(\rho)]^{-1}$ es una función de clase C^∞ en \mathbb{R} . Adicionalmente, se puede comprobar que si $|\rho| \rightarrow \infty$,

$$\begin{aligned} \lim_{|\rho| \rightarrow \infty} H(\rho) &= S_Y^2 - S_{MY_1}^2 / c_2 \\ &= c^* \end{aligned} \quad (3.3.38)$$

con c^* una constante estrictamente positiva como lo prueba el siguiente argumento

$$\begin{aligned} c^* = 0 &\Leftrightarrow S_Y^2 - S_{MY_1}^2 / c_2 = 0 \\ &\Leftrightarrow c_2 S_Y^2 - S_{MY_1}^2 = 0 \\ &\Leftrightarrow c_2 S_Y^2 = S_{MY_1}^2 \\ &\Leftrightarrow [\underline{Z}_1^t (I-P) \underline{Z}_1] [\underline{Y}^t (I-P) \underline{Y}] = [\underline{Z}_1^t (I-P) \underline{Y}]^2 \end{aligned}$$

si se definen

$$\underline{b}_1 = (I-P) \underline{Z}_1$$

$$\underline{b}_2 = (I-P) \underline{Y}$$

y se utiliza nuevamente el hecho de que $(I - P)$ es una matriz idempotente, se obtiene que

$$c^* = 0 \Leftrightarrow (\underline{b}_1^t \underline{b}_1) (\underline{b}_2^t \underline{b}_2) = (\underline{b}_1^t \underline{b}_2)^2$$

de donde mediante una aplicación de la desigualdad de Schwarz se sigue que

$$c^{\circ} = 0 \Leftrightarrow \exists \alpha \in \mathbb{R} \text{ tal que } \underline{b}_1 = \alpha \underline{b}_2$$

Sin embargo,

$$\begin{aligned} \underline{b}_1 &= \alpha \underline{b}_2 && \Leftrightarrow \underline{b}_1 - \alpha \underline{b}_2 = \underline{0} \\ &&& \Leftrightarrow (I-P)\underline{z}_1 - \alpha(I-P)\underline{y} = \underline{0} \\ &&& \Leftrightarrow (I-P)(\underline{z}_1 - \alpha\underline{y}) = \underline{0} \\ &&& \Leftrightarrow (\underline{z}_1 - \alpha\underline{y})^t (I-P)(\underline{z}_1 - \alpha\underline{y}) = 0 \end{aligned}$$

de manera que $c^{\circ} = 0$ si y solo si existe una combinación lineal de las columnas $\underline{z}_0, \underline{z}_1, \dots, \underline{z}_k$ que ajusta con error cero al vector $\underline{z}_1 - \alpha\underline{y}$. Esto es si existe una combinación lineal de los vectores columna de la matriz Z que ajusta con error cero al vector $\alpha\underline{y}$. Si $\alpha = 0$ se obtiene una contradicción puesto que Z es de rango completo. Por otra parte, si $\alpha \neq 0$ como $N > k-1$ simplemente se tiene que

$$p\{c^{\circ} = 0\} = 0 \quad (3.3.39)$$

En conclusión, puesto que si $|\rho| \rightarrow \infty$ $HC(\rho)$ tiende a un límite bien definido y estrictamente positivo, entonces $[HC(\rho)]^{-1}$ tiende, cuando $|\rho| \rightarrow \infty$, a un límite finito y positivo y en consecuencia se puede afirmar que $[HC(\rho)]^{-1}$ es acotada. Finalmente, recurriendo nuevamente a la expresión

$$p(\rho | \underline{y}) \propto p(\rho) [Q(\rho)]^{-1/2} [HC(\rho)]^{-(N+r-k)/2}$$

se tiene que en virtud de que $[HC(\rho)]^{-1}$ es acotada y dado que la función $[Q(\rho)]^{-1/2}$ claramente también tiene esta propiedad, entonces para toda $p(\rho)$ propia $p(\rho | \underline{Y})$ resulta una densidad propia. Más aún, si $p(\rho)$ es impropia pero acotada entonces es claro que si $|\rho|$ es suficientemente grande, existe una constante positiva M tal que

$$p(\rho | \underline{Y}) \leq M [Q(\rho)]^{-1/2}$$

de donde la integrabilidad de $p(\rho | \underline{Y})$ es una consecuencia de la correspondiente integrabilidad de $[Q(\rho)]^{-1/2}$ (Gradshteyn y Ryzhik 1965, página 81). De cualquier forma, el cálculo específico de la integral

$$\int p(\rho) [Q(\rho)]^{-1/2} [HC(\rho)]^{-(N+r-k)/2} d\rho$$

depende de la elección particular que se haga de $p(\rho)$ y en general requiere de algún análisis numérico. Resumiendo, en esta sección se ha establecido el siguiente

Teorema 1. Sea $\underline{Y} \in \mathbb{R}^N$ tal que $\underline{Y} \sim NC(X\theta, \sigma^2 I)$ con X una matriz $(N \times k)$ fija y de rango $k < N$, $\theta \in \mathbb{R}^k$ un vector de coeficientes desconocidos y $\sigma > 0$ también desconocida. Sean además, λ_1 y λ_2 dos vectores fijos, linealmente independientes en \mathbb{R}^k y suponga que bajo la hipótesis $\lambda_2^t \theta \neq 0$, interesa obtener inferencias sobre el parámetro $\rho = \lambda_1^t \theta / \lambda_2^t \theta$. Si se asigna una densidad inicial conjunta para (θ, σ) de la forma

$$p(\theta, \sigma) \propto p(\rho) \sigma^{-r} / |\lambda_2^t \theta|$$

donde $p(\rho)$ es una función acotada de ρ , entonces la correspondiente densidad marginal para el parámetro de interés ρ es propia y está dada por

$$p(\rho | \underline{Y}) \propto p(\rho) [Q(\rho)]^{-1/2} [H(\rho)]^{-(N+r-k)/2}$$

con $Q(\rho)$ y $H(\rho)$ definidas en (3.3.15) y (3.3.31) respectivamente.

Observación. Aún cuando no es evidente del enunciado del teorema, las expresiones para las funciones $Q(\rho)$ y $H(\rho)$ prueban que la densidad marginal final $p(\rho | \underline{Y})$ depende de la matriz Z y en particular sugieren que esta densidad podría no ser invariante ante la elección de la matriz L que define la reparametrización que transforma $\underline{\theta}$ en $\underline{\beta}$. En las siguientes secciones se examinará esta cuestión así como la relación que guarda $p(\rho | \underline{Y})$ con las regiones de confianza que se obtienen utilizando el Teorema de Fieller.

3.4 Invarianza ante reparametrizaciones.

De la expresión (3.3.34) se tiene la siguiente formulación de la densidad final marginal para el parámetro de interés

$$p(\rho | \underline{Y}) \propto p(\rho) [Q(\rho)]^{-1/2} [H(\rho)]^{-(N+r-k)/2}$$

con $Q(\rho) = c_2 \rho^2 + c_1 \rho + c_0$, en donde

$$c_2 = \sum_1^t (I-P) \underline{z}_1$$

$$c_1 = 2 \sum_1^t (I-P) \underline{z}_2$$

$$c_0 = \sum_2^t (I-P) \underline{z}_2$$

y además

$$H(\rho) = (\alpha(\rho) S_Y^2 - \{S_{ZY_2} + \rho S_{ZY_1}\}^2) / \alpha(\rho) \quad (3.4.1)$$

con

$$S_Y^2 = \underline{Y}'(I - P)\underline{Y}$$

$$S_{ZY_2} = \underline{Z}_2'(I - P)\underline{Y}$$

$$S_{ZY_1} = \underline{Z}_1'(I - P)\underline{Y}$$

y

$$P = \underline{Z}_2'(\underline{Z}_2'\underline{Z}_2)^{-1}\underline{Z}_2.$$

Para establecer la invarianza de la densidad final marginal $p(\rho | \underline{Y})$ es conveniente recordar la siguiente

Proposición 1. Los estimadores de mínimos cuadrados $\hat{\beta}_1$ y $\hat{\beta}_2$ para las dos primeras entradas del vector β son invariantes ante la selección particular de la matriz L.

Demostración. El estimador de mínimos cuadrados del vector β está dado por

$$\hat{\beta} = (Z'Z)^{-1}Z'Y$$

pero $Z = X L^{-1}$ y $(Z'Z)^{-1} = L(X'X)^{-1}L'$ (cf. sección 2.2) de donde se sigue que

$$(Z'Z)^{-1}Z'Y = L(X'X)^{-1}X'Y$$

por lo tanto, si e_i es el i -ésimo vector de la base canónica en R^k , entonces

$$\begin{aligned}
 \hat{\beta}_i &= \mathbf{e}_i^t \hat{\theta} \\
 &= \mathbf{e}_i^t [\mathbf{L} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}] \\
 &= (\mathbf{e}_i^t \mathbf{L}) [(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}] \\
 &= \mathbf{1}_i^t \hat{\theta}
 \end{aligned}$$

y en consecuencia, $\hat{\beta}_1 = \lambda_1^t \hat{\theta}$ y $\hat{\beta}_2 = \lambda_2^t \hat{\theta}$ resultan invariantes ante la elección de \mathbf{L} puesto que dependen únicamente de los dos primeros renglones que son siempre los mismos. \square

Un resultado similar puede establecerse para la función $Q(\rho)$ como se verifica en la siguiente

Proposición 2. *El polinomio de grado dos $Q(\rho)$ es invariante ante la selección particular de la matriz \mathbf{L} .*

Demostración. Nuevamente, recordando la definición de la matriz \mathbf{Z} , se tiene que

$$\mathbf{Z} = \mathbf{X} \mathbf{L}^{-1} \quad ; \quad \mathbf{L}^t = (\mathbf{1}_1, \dots, \mathbf{1}_k)$$

y también, en consecuencia

$$(\mathbf{Z}^t \mathbf{Z})^{-1} = \mathbf{L} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{L}^t$$

por tanto si otra vez \mathbf{e}_i es el i -ésimo vector de la base canónica en \mathbb{R}^k , se sigue que

$$\begin{aligned}
 \mathbf{e}_i^t (\mathbf{Z}^t \mathbf{Z})^{-1} \mathbf{e}_j &= \mathbf{e}_i^t [\mathbf{L} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{L}^t] \mathbf{e}_j \\
 &= (\mathbf{e}_i^t \mathbf{L}) (\mathbf{X}^t \mathbf{X})^{-1} (\mathbf{L}^t \mathbf{e}_j)
 \end{aligned}$$

$$= (L^t \underline{e}_i)^t (X^t X)^{-1} (L^t \underline{e}_j)$$

$$= \underline{l}_i^t (X^t X)^{-1} \underline{l}_j$$

de donde

$$\begin{aligned} V_{11} &= \underline{e}_1^t (Z^t Z)^{-1} \underline{e}_1 \\ &= \underline{\lambda}_1^t (X^t X)^{-1} \underline{\lambda}_1 \end{aligned} \quad (3.4.2)$$

$$\begin{aligned} V_{12} &= \underline{e}_1^t (Z^t Z)^{-1} \underline{e}_2 \\ &= \underline{\lambda}_1^t (X^t X)^{-1} \underline{\lambda}_2 \end{aligned} \quad (3.4.3)$$

y

$$\begin{aligned} V_{22} &= \underline{e}_2^t (Z^t Z)^{-1} \underline{e}_2 \\ &= \underline{\lambda}_2^t (X^t X)^{-1} \underline{\lambda}_2 \end{aligned} \quad (3.4.4)$$

son invariantes ante la elección de L puesto que, al igual que $\hat{\beta}_1$ y $\hat{\beta}_2$, solo dependen de $\underline{l}_1 = \underline{\lambda}_1$ y $\underline{l}_2 = \underline{\lambda}_2$. Ahora bien, si se define

$$V_{11} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \quad (3.4.5)$$

resulta que V_{11} es invariante ante la elección de la matriz L. Por otra parte, si se utiliza la partición

$$Z = (Z_1, Z_2) \quad (N \times k)$$

con $Z_1 = (Z_{11}, Z_{12}) \quad (N \times 2)$ y $Z_2 = (Z_{21}, \dots, Z_{2k}) \quad (N \times (k-2))$, se tiene que

$$\begin{aligned}
 Z^t Z &= \begin{bmatrix} Z_1^t \\ Z_2^t \end{bmatrix} (Z_1, Z_2) \\
 &= \begin{bmatrix} Z_1^t Z_1 & Z_1^t Z_2 \\ Z_2^t Z_1 & Z_2^t Z_2 \end{bmatrix} \quad (3.4.6)
 \end{aligned}$$

y como además,

$$(I - P) = (I - Z_2 (Z_2^t Z_2)^{-1} Z_2^t)$$

se sigue que, si se define $V = (v_{ij}) = (Z^t Z)^{-1}$ entonces considerando la partición correspondiente a (3.4.6),

$$V = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

de donde, v_{11} (2×2) admite la representación (Searle 1982, página 269)

$$v_{11} = A^{-1} - A^{-1} B [C^{-1} + B^t A^{-1} B]^{-1} B^t A^{-1}$$

con $A = (Z_1^t Z_1)$, $B = (Z_1^t Z_2)$ y $C = -(Z_2^t Z_2)^{-1}$. En consecuencia (Searle 1982, página 269),

$$\begin{aligned}
 v_{11}^{-1} &= A + B C B^t \\
 &= (Z_1^t Z_1) - (Z_1^t Z_2) (Z_2^t Z_2)^{-1} (Z_2^t Z_1) \\
 &= Z_1^t [I - Z_2 (Z_2^t Z_2)^{-1} Z_2^t] Z_1 \\
 &= Z_1^t (I - P) Z_1 \quad (3.4.7)
 \end{aligned}$$

que por las mismas razones que V_{ii} es invariante ante la elección de L . Ahora, recordando la definición del polinomio $Q(\rho)$, se tiene que

$$Q(\rho) = c_2 \rho^2 + c_1 \rho + c_0$$

con

$$\begin{aligned} c_2 &= \underline{z}_1^t (I-P) \underline{z}_1 \\ &= [Z_{\underline{z}_1}]^t (I-P) [Z_{\underline{z}_1}] \\ &= \underline{e}_1^t \{ Z^t (I-P) Z \} \underline{e}_1 \end{aligned} \quad (3.4.8)$$

$$\begin{aligned} c_1 &= 2 \underline{z}_1^t (I-P) \underline{z}_2 \\ &= 2 [Z_{\underline{z}_1}]^t (I-P) [Z_{\underline{z}_2}] \\ &= 2 \underline{e}_1^t \{ Z^t (I-P) Z \} \underline{e}_2 \end{aligned} \quad (3.4.9)$$

$$\begin{aligned} c_0 &= \underline{z}_2^t (I-P) \underline{z}_2 \\ &= [Z_{\underline{z}_2}]^t (I-P) [Z_{\underline{z}_2}] \\ &= \underline{e}_2^t \{ Z^t (I-P) Z \} \underline{e}_2 \end{aligned} \quad (3.4.10)$$

de manera que $Q(\rho)$ depende de la matriz L únicamente a través de la expresión

$$Z^t (I - P) Z = \begin{bmatrix} Z_1^t \\ Z_2^t \end{bmatrix} (I - P) (Z_1, Z_2)$$

$$\begin{aligned}
 &= \begin{bmatrix} Z_1^t \\ Z_2^t \end{bmatrix} [(I - P)Z_1, (I - P)Z_2] \\
 &= \begin{bmatrix} Z_1^t(I - P)Z_1 & Z_1^t(I - P)Z_2 \\ \cdot & \cdot \\ Z_2^t(I - P)Z_1 & Z_2^t(I - P)Z_2 \end{bmatrix} \quad (3.4.11)
 \end{aligned}$$

sin embargo,

$$\begin{aligned}
 Z_2^t(I - P) &= Z_2^t(I - Z_2(Z_2^t Z_2)^{-1} Z_2^t) \quad (3.4.12) \\
 &= Z_2^t - Z_2^t Z_2 (Z_2^t Z_2)^{-1} Z_2^t \\
 &= 0
 \end{aligned}$$

por lo tanto

$$Z^t(I - P)Z = \begin{bmatrix} Z_1^t(I - P)Z_1 & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \quad (3.4.13)$$

$$= \begin{bmatrix} V^{-1} & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & \cdot \end{bmatrix} \quad (3.4.14)$$

es invariante ante la elección de la matriz L . Finalmente, recurriendo a las expresiones (3.4.8), (3.4.9) y (3.4.10) para los coeficientes de $Q(\rho)$, se sigue la proposición. \square

Existe otro término cuya invarianza ante la reparametrización es oportuno establecer mediante la siguiente

Proposición 3. La suma de Cuadrados $S_Y^2 = Y^t(I - P)Y$ es invariante ante la selección particular de la matriz L .

Demostración. Si se considera el Modelo Completo (reparametrizado)

$$\underline{Y} \sim NC \underline{Z}\underline{\beta}, \sigma^2 \underline{I} \quad (3.4.15)$$

y si ahora se plantea la hipótesis $H: \beta_1 = \beta_2 = 0$, entonces bajo H se obtiene el Modelo Reducido

$$\underline{Y} \sim NC \underline{Z}_2 \underline{\beta}_2, \sigma^2 \underline{I} \quad (3.4.16)$$

con $\underline{\beta}_2^t = (\beta_3, \dots, \beta_k)$, que coincide con el vector $\underline{\gamma}_2$ de la sección 3.3. En términos matriciales la hipótesis H se puede reescribir como

$$H : F \underline{\beta} = \underline{0}$$

con $F^t = (\underline{e}_1, \underline{e}_2)$ ($k \times 2$) y nuevamente utilizando la terminología frecuentista para la prueba de hipótesis se tiene que la suma de cuadrados de la hipótesis está dada por

$$\begin{aligned} SCE_H &= (F \hat{\underline{\beta}})^t [F(Z^t Z)^{-1} F^t]^{-1} (F \hat{\underline{\beta}}) \\ &= (\hat{\beta}_1, \hat{\beta}_2) [F V F^t]^{-1} (\hat{\beta}_1, \hat{\beta}_2)^t \\ &= (\hat{\beta}_1, \hat{\beta}_2) [V_{11}]^{-1} (\hat{\beta}_1, \hat{\beta}_2)^t \end{aligned} \quad (3.4.17)$$

de donde, puesto que en la demostración de la Proposición 2 se ha probado que V_{11}^{-1} es invariante y de la Proposición 1 se sigue que $\hat{\beta}_1$ y $\hat{\beta}_2$ son invariantes, se tiene que SCE_H es invariante. Ahora bien, dado que la suma de cuadrados del Modelo Completo (3.4.15) está dada por

$$SCE_{MC} = \underline{Y}^t (I - Z(Z^t Z)^{-1} Z^t) \underline{Y}$$

de los resultados de la sección 2.2. se tiene que

$$\begin{aligned} \text{SCE}_{\text{MC}} &= \underline{Y}' (I - X(X'X)^{-1}X') \underline{Y} \\ &= (N-k) S^2 \end{aligned} \quad (3.4.18)$$

y también es invariante. Otra aplicación de la identidad

$$\text{SCE}_{\text{MR}} = \text{SCE}_{\text{MC}} + \text{SCE}_{\text{R}}$$

permite concluir que

$$\begin{aligned} \underline{Y}' (I - P) \underline{Y} &= \underline{Y}' (I - Z_2(Z_2'Z_2)^{-1}Z_2') \underline{Y} \\ &= \text{SCE}_{\text{MR}} \end{aligned}$$

es invariante ante la elección de la matriz L, lo cual finaliza la demostración. \square

Hasta aquí, se ha probado que $Q(\rho)$ es invariante ante la elección de L y recordando la expresión

$$\text{HC}(\rho) = [Q(\rho)S_Y^2 - \{S_{ZY2} + \rho S_{ZY1}\}^2] / Q(\rho) \quad (3.4.19)$$

se ha verificado que la magnitud S_Y^2 es también invariante ante L. De esta manera, para probar que $\text{HC}(\rho)$ es invariante basta con probar que $S_{ZY2} + \rho S_{ZY1}$ es invariante. Para tal efecto y utilizando el hecho de que $\hat{\beta}_1$ y $\hat{\beta}_2$ son invariantes, se verificará la siguiente

Proposición 4. S_{ZY1} y S_{ZY2} son invariantes ante la selección particular de la matriz L.

Demostración. En la Proposición 1 se probó que $\hat{\beta}_1$ y $\hat{\beta}_2$ son invariantes en donde $\hat{\beta}_1$ y $\hat{\beta}_2$ son las dos primeras entradas del vector $\hat{\theta}$ que satisface el sistema de ecuaciones normales

$$(Z^t Z) \hat{\theta} = Z^t Y \quad (3.4.20)$$

Si la matriz $Z = (z_1, \dots, z_k)$ ($N \times k$) se expresa de nuevo en forma particionada

$$Z = (Z_1, Z_2) \quad (3.4.21)$$

con $Z_1 = (z_1, z_2)$ ($N \times 2$) y $Z_2 = (z_3, \dots, z_k)$ ($N \times (k-2)$), entonces el sistema (3.4.20) resulta

$$\begin{bmatrix} Z_1^t Z_1 & Z_1^t Z_2 \\ Z_2^t Z_1 & Z_2^t Z_2 \end{bmatrix} \hat{\theta} = \begin{bmatrix} Z_1^t Y \\ Z_2^t Y \end{bmatrix} \quad (3.4.22)$$

Si además $\hat{\theta}$ también se escribe en forma particionada

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \quad (3.4.23)$$

con $\hat{\theta}_1 = (\hat{\beta}_1, \hat{\beta}_2)$ y $\hat{\theta}_2 = (\hat{\beta}_3, \dots, \hat{\beta}_k)$, se tiene que el sistema (3.4.22) se puede escribir como

$$\begin{aligned} (Z_1^t Z_1) \hat{\theta}_1 + (Z_1^t Z_2) \hat{\theta}_2 &= Z_1^t Y \\ (Z_2^t Z_1) \hat{\theta}_1 + (Z_2^t Z_2) \hat{\theta}_2 &= Z_2^t Y \end{aligned} \quad (3.4.24)$$

Si ahora se utiliza la expresión particionada $Z_1 = (z_1, z_2)$ se obtiene

$$\begin{bmatrix} \sum_1^t \sum_1 \\ \vdots \\ \sum_1^t \sum_2 \\ \vdots \\ \sum_2^t \sum_1 \\ \vdots \\ \sum_2^t \sum_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_2 \end{bmatrix} + \begin{bmatrix} \sum_1^t Z_2 \\ \vdots \\ \sum_2^t Z_2 \\ \vdots \\ \sum_2^t Z_2 \end{bmatrix} \hat{\beta}_2 = \begin{bmatrix} \sum_1^t Y \\ \vdots \\ \sum_2^t Y \end{bmatrix} \quad (3.4.25)$$

$$\begin{bmatrix} Z_2^t \sum_1 & Z_2^t \sum_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} + \begin{bmatrix} Z_2^t Z_2 \end{bmatrix} \hat{\beta}_2 = Z_2^t Y$$

o equivalentemente.

$$\begin{aligned} (\sum_1^t \sum_1) \hat{\beta}_1 + (\sum_1^t \sum_2) \hat{\beta}_2 + (\sum_1^t Z_2) \hat{\beta}_2 &= \sum_1^t Y \\ (\sum_2^t \sum_1) \hat{\beta}_1 + (\sum_2^t \sum_2) \hat{\beta}_2 + (\sum_2^t Z_2) \hat{\beta}_2 &= \sum_2^t Y \\ (\sum_2^t \sum_1) \hat{\beta}_1 + (\sum_2^t \sum_2) \hat{\beta}_2 + (\sum_2^t Z_2) \hat{\beta}_2 &= Z_2^t Y \end{aligned}$$

... (3.4.26)

de donde se tiene que

$$\hat{\beta}_1 = |Z^t Z|^{-1} \begin{bmatrix} \sum_1^t Y & \sum_1^t \sum_2 & \sum_1^t Z_2 \\ \sum_2^t Y & \sum_2^t \sum_2 & \sum_2^t Z_2 \\ Z_2^t Y & Z_2^t \sum_2 & Z_2^t Z_2 \end{bmatrix} \quad (3.4.27)$$

si ahora se define $Z_0 = (\sum_1, \sum_2, Z_2)$ es claro que

$$\begin{bmatrix} \sum_1^t Y & \sum_1^t \sum_2 & \sum_1^t Z_2 \\ \sum_2^t Y & \sum_2^t \sum_2 & \sum_2^t Z_2 \\ Z_2^t Y & Z_2^t \sum_2 & Z_2^t Z_2 \end{bmatrix} = \begin{bmatrix} \sum_1^t Y & \vdots & \sum_1^t Z_0 \\ \sum_2^t Y & \vdots & \sum_2^t Z_0 \\ Z_0^t Y & \vdots & Z_0^t Z_0 \end{bmatrix} \quad (3.4.28)$$

de manera que el numerador en (3.4.27) puede calcularse (Morrison 1976, página 68) como sigue

$$\begin{aligned} \begin{bmatrix} \sum_1^t Y & \sum_1^t Z_0 \\ \sum_2^t Y & \sum_2^t Z_0 \end{bmatrix} &= |Z_0^t Z_0| \left| \sum_1^t Y - \sum_1^t Z_0 (Z_0^t Z_0)^{-1} \sum_1^t Y \right| \\ &= |Z_0^t Z_0| \left\{ \sum_1^t (I - Z_0 (Z_0^t Z_0)^{-1} Z_0^t) Y \right\}. \end{aligned}$$

Con un argumento similar se obtiene

$$\begin{aligned} |Z_0^t Z_0| &= \begin{bmatrix} \sum_2^t Z_2 & \sum_2^t Z_2 \\ \sum_2^t Z_2 & \sum_2^t Z_2 \end{bmatrix} \\ &= |Z_2^t Z_2| \left| \sum_2^t Z_2 - \sum_2^t Z_2 (Z_2^t Z_2)^{-1} \sum_2^t Z_2 \right| \\ &= |Z_2^t Z_2| \left\{ \sum_2^t (I - Z_2 (Z_2^t Z_2)^{-1} Z_2^t) Z_2 \right\} \\ &= |Z_2^t Z_2| \left\{ \sum_2^t (I - P) Z_2 \right\} \\ &= |Z_2^t Z_2| c_0 \end{aligned} \tag{3.4.20}$$

Por otra parte, si en correspondencia con la partición de $Z_0^t Z_0$ se define $D = (Z_0^t Z_0)^{-1}$ como

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

entonces (Morrison 1976, página 66).

$$\begin{aligned} D_{11} &= \left[\sum_2^t Z_2 - \sum_2^t Z_2 (Z_2^t Z_2)^{-1} \sum_2^t Z_2 \right]^{-1} \\ &= \left[\sum_2^t (I - Z_2 (Z_2^t Z_2)^{-1} Z_2^t) Z_2 \right]^{-1} \\ &= \left[\sum_2^t (I - P) Z_2 \right]^{-1} \end{aligned}$$

$$D_{12} = - [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t Z_2 (Z_2^t Z_2)^{-1}$$

$$D_{21} = (D_{12})^t$$

$$= - (Z_2^t Z_2)^{-1} Z_2^t \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1}$$

$$D_{22} = (Z_2^t Z_2)^{-1} + (Z_2^t Z_2)^{-1} Z_2^t \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t Z_2 (Z_2^t Z_2)^{-1}$$

de manera que

$$\begin{aligned} Z_0 (Z_0^t Z_0)^{-1} Z_0^t &= \begin{bmatrix} \underline{z}_2 & Z_2 \end{bmatrix} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} \underline{z}_2^t \\ Z_2^t \end{bmatrix} \\ &= \underline{z}_2 D_{11} \underline{z}_2^t + \underline{z}_2 D_{12} Z_2^t + Z_2 D_{21} \underline{z}_2^t + Z_2 D_{22} Z_2^t \\ &= \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t \\ &\quad - \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t Z_2 (Z_2^t Z_2)^{-1} Z_2^t \\ &\quad - Z_2 (Z_2^t Z_2)^{-1} Z_2^t \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t \\ &\quad + Z_2 (Z_2^t Z_2)^{-1} Z_2^t \\ &\quad + Z_2 (Z_2^t Z_2)^{-1} Z_2^t \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t Z_2 (Z_2^t Z_2)^{-1} Z_2^t \\ &= \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t \\ &\quad - \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t P \\ &\quad - P \underline{z}_2 [\underline{z}_2^t (I - P) \underline{z}_2]^{-1} \underline{z}_2^t \end{aligned}$$

$$\begin{aligned}
 &+ P + P \Xi_2 [\Xi_2^t (I-P) \Xi_2]^{-1} \Xi_2^t P \\
 &= P + (I-P) \Xi_2 [\Xi_2^t (I-P) \Xi_2]^{-1} \Xi_2^t (I-P) \quad (3.4.30)
 \end{aligned}$$

Por lo tanto,

$$\begin{aligned}
 \Xi_1^t [I - Z_0 (Z_0^t Z_0)^{-1} Z_0^t] Y &= \Xi_1^t (I-P) Y \\
 &- \Xi_1^t (I-P) \Xi_2 [\Xi_2^t (I-P) \Xi_2]^{-1} \Xi_2^t (I-P) Y \\
 &= S_{SY1} - (c_1 / 2) (c_0)^{-1} S_{SY2} \\
 &= S_{SY1} - c_1 S_{SY2} / (2c_0) \quad (3.4.31)
 \end{aligned}$$

y en consecuencia

$$\hat{\beta}_1 = |Z^t Z|^{-1} |Z_1^t Z_1| c_0 \{ S_{SY1} - c_1 S_{SY2} / (2c_0) \} \quad (3.4.32)$$

Adicionalmente, utilizando la partición (3.4.6) de la matriz $Z^t Z$, se tiene que en forma análoga a (3.4.29)

$$\begin{aligned}
 |Z^t Z| &= |Z_2^t Z_2| |Z_1^t Z_1 - (Z_1^t Z_2) (Z_2^t Z_2)^{-1} (Z_2^t Z_1)| \\
 &= |Z_2^t Z_2| |Z_1^t [I - Z_2 (Z_2^t Z_2)^{-1} Z_2^t] Z_1| \\
 &= |Z_2^t Z_2| |Z_1^t (I - P) Z_1|
 \end{aligned}$$

y utilizando (3.4.7),

$$|Z^t Z| = |Z_2^t Z_2| |V_{11}|^{-1} \quad (3.4.33)$$

Por último, entonces

$$\hat{\beta}_1 = |V_{11}| c_0 \{ S_{ZY1} - c_1 S_{ZY2} / (2c_0) \}$$

o bien

$$\hat{\beta}_1 = |V_{11}| c_0 S_{ZY1} - |V_{11}| K c_1 / 2 S_{ZY2} \quad (3.4.34)$$

Del mismo sistema de ecuaciones (3.4.20) se tiene

$$\hat{\beta}_2 = |Z^t Z|^{-1} \begin{bmatrix} \sum_1^t \varepsilon_1 \\ \sum_2^t \varepsilon_1 \\ Z_2^t \varepsilon_1 \end{bmatrix} \begin{bmatrix} \sum_1^t Y \\ \sum_2^t Y \\ Z_2^t Y \end{bmatrix} \begin{bmatrix} \sum_1^t Z_2 \\ \sum_2^t Z_2 \\ Z_2^t Z_2 \end{bmatrix} \quad (3.4.35)$$

$$= - |Z^t Z|^{-1} \begin{bmatrix} \sum_1^t Y \\ \sum_2^t Y \\ Z_2^t Y \end{bmatrix} \begin{bmatrix} \sum_1^t \varepsilon_1 \\ \sum_2^t \varepsilon_1 \\ Z_2^t \varepsilon_1 \end{bmatrix} \begin{bmatrix} \sum_1^t Z_2 \\ \sum_2^t Z_2 \\ Z_2^t Z_2 \end{bmatrix}$$

$$= |Z^t Z|^{-1} \begin{bmatrix} \sum_2^t Y \\ \sum_1^t Y \\ Z_2^t Y \end{bmatrix} \begin{bmatrix} \sum_2^t \varepsilon_1 \\ \sum_1^t \varepsilon_1 \\ Z_2^t \varepsilon_1 \end{bmatrix} \begin{bmatrix} \sum_2^t Z_2 \\ \sum_1^t Z_2 \\ Z_2^t Z_2 \end{bmatrix} \quad (3.4.36)$$

$$= |Z^t Z|^{-1} \begin{bmatrix} \sum_2^t Y & \dots & \sum_2^t Z_2 \\ Z_2^t Y & \dots & Z_2^t Z_2 \end{bmatrix} \quad (3.4.37)$$

con $Z_+ = (\sum_2, Z_2)$. Procediendo en forma análoga al caso de $\hat{\beta}_1$, se tiene que

$$\begin{bmatrix} \sum_2^t Y & \dots & \sum_2^t Z_2 \\ Z_2^t Y & \dots & Z_2^t Z_2 \end{bmatrix} = |Z_+^t Z_+| \left| \sum_2^t Y - \sum_2^t Z_2 (Z_+^t Z_+)^{-1} Z_+^t Y \right|$$

$$= |Z_1' Z_1| \{ \underline{z}_2' [I - Z_1 (Z_1' Z_1)^{-1} Z_1'] \underline{y} \} \quad (3.4.38)$$

Ahora bien,

$$\begin{aligned} |Z_1' Z_1| &= |Z_2' Z_2| \{ \underline{z}_1' \underline{z}_1 - \underline{z}_1' Z_2 (Z_2' Z_2)^{-1} Z_2' \underline{z}_1 \} \\ &= |Z_2' Z_2| \{ \underline{z}_1' [I - Z_2 (Z_2' Z_2)^{-1} Z_2'] \underline{z}_1 \} \\ &= |Z_2' Z_2| \{ \underline{z}_1' [I - P] \underline{z}_1 \} \\ &= |Z_2' Z_2| c_2 \end{aligned} \quad (3.4.39)$$

y que dado que es posible verificar que

$$Z_1 (Z_1' Z_1)^{-1} Z_1' = P + (I-P) \underline{z}_1 [\underline{z}_1' (I-P) \underline{z}_1]^{-1} \underline{z}_1' (I-P)$$

se sigue que

$$\begin{aligned} \underline{z}_2' [I - Z_1 (Z_1' Z_1)^{-1} Z_1'] \underline{y} &= \underline{z}_2' (I-P) \underline{y} \\ &= \underline{z}_2' (I-P) \underline{z}_1 [\underline{z}_1' (I-P) \underline{z}_1]^{-1} \underline{z}_2' (I-P) \underline{y} \\ &= S_{Z_2 Y_2} - (c_1/2)(c_2)^{-1} S_{Z_2 Y_1} \\ &= S_{Z_2 Y_2} - c_1 S_{Z_2 Y_1} / (2c_2) \end{aligned} \quad (3.4.40)$$

y por tanto, utilizando (3.4.37), (3.4.38), (3.4.39) y (3.4.40) se puede comprobar que

$$\begin{aligned} \hat{\beta}_2 &= |Z_1' Z_1|^{-1} |Z_2' Z_2| c_2 \{ S_{Z_2 Y_2} - c_1 S_{Z_2 Y_1} / (2c_2) \} \\ &= |V_{11}| c_2 \{ S_{Z_2 Y_2} - c_1 S_{Z_2 Y_1} / (2c_2) \} \end{aligned}$$

$$= |V_{11}| c_2 S_{ZV2} - |V_{11}| (c_1/2) S_{ZV1} \quad (3.4.41)$$

Como resultado de (3.4.34) y (3.4.41) se obtiene el sistema de ecuaciones lineales, en S_{ZV1} y S_{ZV2} :

$$\begin{aligned} |V_{11}| c_0 S_{ZV1} - |V_{11}| (c_1/2) S_{ZV2} &= \hat{\beta}_1 \\ |V_{11}| c_2 S_{ZV2} - |V_{11}| (c_1/2) S_{ZV1} &= \hat{\beta}_2 \end{aligned} \quad (3.4.42)$$

En forma matricial, el sistema (3.4.42) puede escribirse como sigue

$$|V_{11}| \begin{bmatrix} c_0 & & & -c_1/2 \\ & c_2 & & \\ & & c_2 & \\ -c_1/2 & & & c_0 \end{bmatrix} \begin{bmatrix} S_{ZV1} \\ S_{ZV2} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad (3.4.43)$$

y puesto que la inversa de la matriz del sistema está dada por

$$|V_{11}|^{-1} \Delta \begin{bmatrix} c_2 & & c_1/2 \\ & c_2 & \\ c_1/2 & & c_0 \end{bmatrix}$$

con $\Delta = (c_0 c_2 - c_1^2/4)^{-1}$, se tiene que

$$\begin{bmatrix} S_{ZV1} \\ S_{ZV2} \end{bmatrix} = |V_{11}|^{-1} \Delta \begin{bmatrix} c_2 & & c_1/2 \\ & c_2 & \\ c_1/2 & & c_0 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

o equivalentemente

$$\begin{aligned} S_{ZV1} &= |V_{11}|^{-1} \Delta c_2 \hat{\beta}_1 + |V_{11}|^{-1} \Delta (c_1/2) \hat{\beta}_2 \\ S_{ZV2} &= |V_{11}|^{-1} \Delta (c_1/2) \hat{\beta}_1 + |V_{11}|^{-1} \Delta c_0 \hat{\beta}_2 \end{aligned} \quad (3.4.44)$$

de donde se tiene que S_{ZV1} y S_{ZV2} dependen de L únicamente a

través de la matriz V_{11} y las magnitudes c_0 , c_1 , c_2 , $\hat{\beta}_1$ y $\hat{\beta}_2$. Por lo tanto, S_{ZY_1} y S_{ZY_2} son invariantes y la proposición queda establecida. \square

Es posible ahora, formular el siguiente teorema que da título a esta sección.

Teorema 1. La densidad marginal final del parámetro de interés ρ .

$$p(\rho | \underline{Y}) \propto p(\rho) [Q(\rho)]^{-1/2} [K(\rho)]^{-(n+r-k)}$$

es invariante ante la matriz de reparametrización L.

Demostración. La densidad marginal inicial $p(\rho)$ es independiente de L y por la Proposición 2, $Q(\rho)$ es invariante. Ahora puesto que

$$K(\rho) = [Q(\rho)S_Y^2 - \{S_{ZY_2} + \rho S_{ZY_1}\}^2] / Q(\rho)$$

y dado que por las Proposiciones 3 y 4, S_Y^2 , S_{ZY_1} y S_{ZY_2} son invariantes, se tiene que $K(\rho)$ es también invariante y el teorema se sigue. \square

En la siguiente sección las expresiones desarrolladas para los componentes de $p(\rho | \underline{Y})$ serán utilizadas para establecer la relación que guardan los resultados que se obtienen con esta densidad final y las inferencias frecuentistas.

3.5 Relación con la regiones de Fieller.

Para establecer la relación entre los resultados bayesianos y frecuentistas es necesario hacer aún más explícitas algunas de las expresiones desarrolladas en la sección 3.4.

Proposición 1. El polinomio $Q(\rho)$ puede ser reexpresado como

$$Q(\rho) = |V_{11}|^{-1} \{ v_{12} \rho^2 - 2 v_{12} \rho + v_{11} \}$$

en donde

$$V_{11} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

es la matriz definida en (3.4.5).

Demostración. De las expresiones (3.4.8), (3.4.9) y (3.4.10) para los coeficientes de $Q(\rho)$ se tiene que

$$c_2 = \underline{e}_1^t \{ Z^t(I-P)Z \} \underline{e}_1$$

$$c_2 = 2\underline{e}_1^t \{ Z^t(I-P)Z \} \underline{e}_2$$

$$c_0 = \underline{e}_2^t \{ Z^t(I-P)Z \} \underline{e}_2$$

Ahora bien, de acuerdo con (3.4.14)

$$Z^t(I-P)Z = \begin{bmatrix} v_{11}^{-1} & 0 \\ \cdot & \cdot \\ 0 & 0 \end{bmatrix}$$

de manera que

$$\begin{aligned}
 c_2 &= \mathbf{e}_1^t \begin{bmatrix} \cdot & v_{11}^{-1} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \mathbf{e}_1 \\
 &= (1, 0) v_{11}^{-1} (1, 0)^t
 \end{aligned} \tag{3.5.1}$$

$$\begin{aligned}
 c_1 &= 2 \mathbf{e}_1^t \begin{bmatrix} \cdot & v_{11}^{-1} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \mathbf{e}_2 \\
 &= (1, 0) v_{11}^{-1} (0, 1)^t
 \end{aligned} \tag{3.5.2}$$

$$\begin{aligned}
 c_0 &= \mathbf{e}_2^t \begin{bmatrix} \cdot & v_{11}^{-1} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \mathbf{e}_2 \\
 &= (0, 1) v_{11}^{-1} (0, 1)^t
 \end{aligned} \tag{3.5.3}$$

y puesto que directamente se puede verificar que

$$v_{11}^{-1} = |V_{11}|^{-1} \begin{bmatrix} v_{22} & -v_{12} \\ -v_{12} & v_{11} \end{bmatrix} \tag{3.5.4}$$

se tiene que

$$c_2 = v_{22} |V_{11}|^{-1} \tag{3.5.5}$$

$$c_1 = -2v_{12} |V_{11}|^{-1} \tag{3.5.6}$$

$$c_0 = v_{11} |V_{11}|^{-1} \tag{3.5.7}$$

lo que permite escribir

$$Q(\rho) = |V_{11}|^{-1} \{ v_{22} \rho^2 - 2v_{12} \rho + v_{11} \} \tag{3.5.8}$$

lo que finaliza la demostración. \square

Proposición 2. La magnitud $S_Y^2 = \underline{y}'(I-P)\underline{y}$ satisface la siguiente identidad

$$\underline{y}'(I-P)\underline{y} = (N-k)S^2 + \{ c_2 \hat{\beta}_1^2 + c_1 \hat{\beta}_1 \hat{\beta}_2 + c_0 \hat{\beta}_2^2 \}$$

con S^2 definida en (3.4.18).

Demostración. De acuerdo a la Proposición 3.

$$\underline{y}'(I-P)\underline{y} = (N-k)S^2 + \langle \hat{\beta}_1, \hat{\beta}_2 \rangle [V_{11}]^{-1} \langle \hat{\beta}_1, \hat{\beta}_2 \rangle'$$

y de acuerdo con (3.5.3)

$$\begin{aligned} \langle \hat{\beta}_1, \hat{\beta}_2 \rangle [V_{11}]^{-1} \langle \hat{\beta}_1, \hat{\beta}_2 \rangle' &= |V_{11}|^{-1} \langle \hat{\beta}_1, \hat{\beta}_2 \rangle \begin{bmatrix} v_{22} & -v_{12} \\ -v_{12} & v_{11} \end{bmatrix} \langle \hat{\beta}_1, \hat{\beta}_2 \rangle' \\ &= |V_{11}|^{-1} \{ v_{22} \hat{\beta}_1^2 - 2v_{12} \hat{\beta}_1 \hat{\beta}_2 + v_{11} \hat{\beta}_2^2 \} \\ &= c_2 \hat{\beta}_1^2 + c_1 \hat{\beta}_1 \hat{\beta}_2 + c_0 \hat{\beta}_2^2 \end{aligned}$$

con lo que se prueba el resultado. \square

Finalmente, antes de establecer el teorema que ilustra la relación de $p(p | \underline{y})$ con las regiones de Fieller, basta considerar la siguiente

Proposición 3. S_{ZY1} y S_{ZY2} satisfacen las identidades

$$S_{ZY1} = c_2 \hat{\beta}_1 + (c_1/2) \hat{\beta}_2$$

$$S_{ZY2} = (c_1/2) \hat{\beta}_1 + c_0 \hat{\beta}_2.$$

Demostración. Basta con observar que

$$\begin{aligned}
 \Delta &= \{ c_0 c_2 - c_1^2 / 4 \}^{-1} \\
 &= \{ v_{11} v_{22} / |V_{11}|^2 - 4v_{12}^2 / (4|V_{11}|^2) \}^{-1} \\
 &= |V_{11}|^2 \{ v_{11} v_{22} - v_{12}^2 \}^{-1} \\
 &= |V_{11}|^2 |V_{11}|^{-1} \\
 &= |V_{11}|
 \end{aligned}$$

de donde, susstituyendo en (3.4.44), se obtiene la proposición. \square

El resultado principal de esta sección se resume en el siguiente

Teorema 1. *La densidad marginal final del parámetro de interés ρ , $K(\rho | Y)$, puede expresarse como sigue*

$$K(\rho | Y) \propto K(\rho) [Q(\rho)]^{-1/2} [\eta + h(\rho)]^{-(\eta+1)/2}$$

en donde $\eta = (N-k)$ y

$$h(\rho) = (\hat{\beta}_1 - \rho \hat{\beta}_2)^2 / [S^2 \{ v_{22} \rho^2 - 2 v_{12} \rho + v_{11} \}]$$

es la variable pivotal de Fieller.

Demostración. Para establecer el resultado es suficiente con probar que

$$H(\rho) \propto [\eta + h(\rho)]$$

ahora bien, de (3.3.31) se tiene que

$$H(\rho) = S_Y^2 - [S_{ZY2} + \rho S_{ZY1}]^2 / Q(\rho)$$

y por la Proposición 2,

$$S_Y^2 = \eta S^2 + \{ c_2 \hat{\beta}_1^2 + c_1 \hat{\beta}_1 \hat{\beta}_2 + c_0 \hat{\beta}_2^2 \}$$

en consecuencia,

$$\begin{aligned} H(\rho) &= \eta S^2 + \{ \{ c_2 \hat{\beta}_1^2 + c_1 \hat{\beta}_1 \hat{\beta}_2 + c_0 \hat{\beta}_2^2 \} Q(\rho) \} / Q(\rho) \\ &\quad - \{ \{ S_{ZY2} + \rho S_{ZY1} \}^2 \} / Q(\rho) \end{aligned}$$

Por otra parte de la Proposición 3, se sigue que

$$\begin{aligned} [S_{ZY2} + \rho S_{ZY1}]^2 &= \{ (c_1 \hat{\beta}_1 / 2 + c_0 \hat{\beta}_2) + \rho (c_2 \hat{\beta}_1 + c_1 \hat{\beta}_2 / 2) \}^2 \\ &= \{ c_2 \hat{\beta}_1 + c_1 \hat{\beta}_2 / 2 \}^2 \rho^2 \\ &\quad + 2 \{ c_1 \hat{\beta}_1 / 2 + c_0 \hat{\beta}_2 \} \{ c_2 \hat{\beta}_1 + c_1 \hat{\beta}_2 / 2 \} \rho \\ &\quad + \{ c_1 \hat{\beta}_1 / 2 + c_0 \hat{\beta}_2 \}^2 \end{aligned}$$

por otra parte, se puede escribir

$$\begin{aligned} \{ c_2 \hat{\beta}_1^2 + c_1 \hat{\beta}_1 \hat{\beta}_2 + c_0 \hat{\beta}_2^2 \} Q(\rho) &= \{ c_2^2 \hat{\beta}_1^2 + c_1 c_2 \hat{\beta}_1 \hat{\beta}_2 + c_0 c_2 \hat{\beta}_2^2 \} \rho^2 \\ &= \{ c_2 c_1 \hat{\beta}_1^2 + c_1^2 \hat{\beta}_1 \hat{\beta}_2 + c_0 c_1 \hat{\beta}_2^2 \} \rho \\ &= \{ c_2 c_0 \hat{\beta}_1^2 + c_1 c_0 \hat{\beta}_1 \hat{\beta}_2 + c_0^2 \hat{\beta}_2^2 \} \end{aligned}$$

Por lo tanto,

$$(c_2 \hat{\beta}_1^2 + c_1 \hat{\beta}_1 \hat{\beta}_2 + c_0 \hat{\beta}_2^2) K(\rho) + (S_{\text{EVI}2} + \rho S_{\text{EVI}1})^2 = d_2 \rho^2 + d_1 \rho + d_0$$

con

$$\begin{aligned} d_2 &= [c_2^2 \hat{\beta}_1^2 + c_1 c_2 \hat{\beta}_1 \hat{\beta}_2 + c_0 c_2 \hat{\beta}_2^2] - [c_2 \hat{\beta}_1 + c_1 \hat{\beta}_2 / 2]^2 \\ &= c_2^2 \hat{\beta}_1^2 + c_1 c_2 \hat{\beta}_1 \hat{\beta}_2 + c_0 c_2 \hat{\beta}_2^2 \\ &\quad - c_2^2 \hat{\beta}_1^2 - c_1 c_2 \hat{\beta}_1 \hat{\beta}_2 - c_1^2 \hat{\beta}_2^2 / 4 \\ &= (c_0 c_2 - c_1^2 / 4) \hat{\beta}_2^2 \\ &= |V_{11}|^{-1} \hat{\beta}_2^2 \end{aligned} \quad (3.5.9)$$

$$\begin{aligned} d_1 &= [c_2 c_1 \hat{\beta}_1^2 + c_1^2 \hat{\beta}_1 \hat{\beta}_2 + c_0 c_1 \hat{\beta}_2^2] \\ &\quad - 2[c_1 \hat{\beta}_1 / 2 + c_0 \hat{\beta}_2][c_2 \hat{\beta}_1 + c_1 \hat{\beta}_2 / 2] \\ &= c_2 c_1 \hat{\beta}_1^2 + c_1^2 \hat{\beta}_1 \hat{\beta}_2 + c_0 c_1 \hat{\beta}_2^2 \\ &\quad - c_2 c_1 \hat{\beta}_1^2 - (c_1^2 / 2) + 2c_0 c_1 \hat{\beta}_1 \hat{\beta}_2 - c_0 c_1 \hat{\beta}_2^2 \\ &= \{ c_1^2 - (c_1^2 / 2) - 2c_0 c_1 \} \hat{\beta}_1 \hat{\beta}_2 \\ &= -2(c_0 c_2 - c_1^2 / 4) \hat{\beta}_1 \hat{\beta}_2 \\ &= -2 |V_{11}|^{-1} \hat{\beta}_1 \hat{\beta}_2 \end{aligned} \quad (3.5.10)$$

$$\begin{aligned} d_0 &= [c_2 c_0 \hat{\beta}_1^2 + c_1 c_0 \hat{\beta}_1 \hat{\beta}_2 + c_0^2 \hat{\beta}_2^2] - [c_1 \hat{\beta}_1 / 2 + c_0 \hat{\beta}_2]^2 \\ &= c_2 c_0 \hat{\beta}_1^2 + c_1 c_0 \hat{\beta}_1 \hat{\beta}_2 + c_0^2 \hat{\beta}_2^2 \\ &\quad - c_1^2 \hat{\beta}_1^2 / 4 - c_1 c_0 \hat{\beta}_1 \hat{\beta}_2 - c_0^2 \hat{\beta}_2^2 \end{aligned}$$

$$\begin{aligned}
 &= (c_2 c_0 - c_1^2 / 4) \hat{\beta}_1^2 \\
 &= |V_{11}|^{-1} \hat{\beta}_1^2 \quad (3.5.11)
 \end{aligned}$$

pero entonces

$$\begin{aligned}
 d_2 \rho^2 + d_1 \rho + d_0 &= |V_{11}|^{-1} \hat{\beta}_2^2 \rho^2 - 2 |V_{11}|^{-1} \hat{\beta}_1 \hat{\beta}_2 \rho \\
 &\quad + |V_{11}|^{-1} \hat{\beta}_1^2 \\
 &= |V_{11}|^{-1} \{ \hat{\beta}_1^2 - 2 \hat{\beta}_1 \hat{\beta}_2 \rho + \hat{\beta}_2^2 \rho^2 \} \\
 &= |V_{11}|^{-1} (\hat{\beta}_1 - \rho \hat{\beta}_2)^2
 \end{aligned}$$

y por lo tanto

$$\begin{aligned}
 h(\rho) &= \eta S^2 + [|V_{11}|^{-1} (\hat{\beta}_1 - \rho \hat{\beta}_2)^2] / Q(\rho) \\
 &= \eta S^2 + [|V_{11}|^{-1} (\hat{\beta}_1 - \rho \hat{\beta}_2)^2] / [|V_{11}|^{-1} \{ v_{22} \rho^2 - 2v_{12} \rho + v_{11} \}] \\
 &= \eta S^2 + (\hat{\beta}_1 - \rho \hat{\beta}_2)^2 / \{ v_{22} \rho^2 - 2v_{12} \rho + v_{11} \} \\
 &= S^2 \{ \eta + h(\rho) \}
 \end{aligned}$$

de manera que

$$\begin{aligned}
 p(\rho | \underline{y}) &\propto p(\rho) [Q(\rho)]^{-1/2} [h(\rho)]^{-(\eta+1)/2} \\
 &= p(\rho) [Q(\rho)]^{-1/2} [S^2 \{ \eta + h(\rho) \}]^{-(\eta+1)/2} \\
 &\propto p(\rho) [Q(\rho)]^{-1/2} [\eta + h(\rho)]^{-(\eta+1)/2}
 \end{aligned}$$

lo que concluye la demostración. \square

La relación que establece el Teorema 1 entre la densidad marginal final $p(\rho | \underline{Y})$ y la variable pivotal $h(\rho)$ puede examinarse con mayor detalle. En primer lugar, es claro que $p(\rho | \underline{Y})$ depende del vector de observaciones únicamente a través de $h(\rho)$. Este hecho podría interpretarse en el sentido de que si se utiliza una densidad inicial conjunta que pertenezca a la familia propuesta en la sección 3.2, entonces la información experimental relevante respecto al parámetro de interés ρ está contenida en la función $h(\rho)$, la variable pivotal de Fieller. Esta interpretación a su vez sugiere que sólo bajo ciertas condiciones específicas la variable pivotal de Fieller puede considerarse un resumen apropiado de la información que el experimento proporciona respecto a ρ .

En otro orden de ideas, se ha mencionado con anterioridad que en contraparte a las llamadas regiones de confianza frecuentistas, desde una perspectiva bayesiana es posible obtener regiones de máxima densidad para el parámetro de interés. También es cierto y bien conocido el hecho de que las regiones de máxima densidad que se obtienen cuando se utilizan algunas densidades iniciales no-informativas coinciden numéricamente con las correspondientes regiones de confianza frecuentistas (cf. Box y Tiao 1973, sección 2.2). En estas condiciones resulta de interés caracterizar las regiones de máxima densidad que se pueden obtener a partir de $p(\rho | \underline{Y})$ y establecer la relación que guardan con las correspondientes regiones de confianza de Fieller.

De la expresión (2.4.2) se tiene que toda región de confianza basada en el procedimiento de Fieller es de la forma

$$R_{(1-\alpha)}^* = \left\{ \rho \in R \mid h(\rho) \leq t^2 \right\} \quad (3.5.12)$$

con t el cuantil de orden $(1-\alpha/2)$ de una distribución t de Student con $(N-k)$ grados de libertad. Es conveniente observar que

$$h(\rho) \leq t^2$$

si y sólo si

$$[\eta + h(\rho)]^{(\eta+r)} \leq c_0 \quad (3.5.13)$$

con $c_0 = [\eta + t^2]^{(\eta+r)}$. De esta forma,

$$R_{(1-\alpha)}^* = \left\{ \rho \in \mathbb{R} \mid [\eta + h(\rho)]^{(\eta+r)} \leq c_0 \right\} \quad (3.5.14)$$

en donde, puesto que t^2 es una función monótona decreciente de α , c_0 resulta una función monótona creciente - no acotada - de $(1-\alpha)$, el tamaño de la región $R_{(1-\alpha)}^*$.

Mientras tanto, una región de máxima densidad final para ρ de tamaño $(1-\alpha)$ es, por definición, una región $R_{(1-\alpha)}$ tal que

$$i) \quad p[\rho \in R_{(1-\alpha)} \mid \underline{Y}] = 1-\alpha$$

$$ii) \quad \text{Para todo par } \rho_1 \in R_{(1-\alpha)} \text{ y } \rho \notin R_{(1-\alpha)}, \text{ se tiene que } p(\rho_1 \mid \underline{Y}) \geq p(\rho \mid \underline{Y}).$$

Es fácil comprobar que la condición ii) es equivalente a

$$ii') \quad \text{Para alguna constante } c \geq 0,$$

$$R_{(1-\alpha)} = \left\{ \rho \in \mathbb{R} \mid p(\rho \mid \underline{Y}) \geq c \right\}.$$

Utilizando ii') y el resultado del Teorema 1, se puede

comprobar que toda región de máxima densidad final para ρ admite la siguiente representación

$$R_{(1-\alpha)} = \left\{ \rho \in \mathbb{R} \mid k_0 p(\rho) [Q(\rho)]^{-1/2} [\eta + h(\rho)]^{-(\eta+r)} \geq c \right\}$$

en donde $k_0 \geq 0$ es una constante tal que $\int p(\rho | \underline{Y}) d\rho = 1$. Ahora bien,

$$k_0 p(\rho) [Q(\rho)]^{-1/2} [\eta + h(\rho)]^{-(\eta+r)/2} \geq c \quad (3.5.15)$$

si y sólo si

$$p(\rho) [Q(\rho)]^{-1/2} [\eta + h(\rho)]^{-(\eta+r)/2} \geq c' \quad (3.5.16)$$

con $c' = c / k_0$. Más aún, la relación (3.5.16) es equivalente a la condición

$$Q(\rho) [\eta + h(\rho)]^{(\eta+r)} \leq c'' \quad (3.5.17)$$

con $Q(\rho) = Q(\rho) / [p(\rho)]^2$ y $c'' = [c']^{-2}$. Si se analiza la condición ii') es fácil concluir que la constante $c \geq 0$ es una función monótona decreciente del tamaño $(1-\alpha)$ de la región. Aún más, es sencillo comprobar que $c \rightarrow 0$ si $(1-\alpha) \rightarrow 1$. En consecuencia, $c' \rightarrow 0$ si $(1-\alpha) \rightarrow 1$ y $c'' \rightarrow \infty$ si $(1-\alpha) \rightarrow 1$. En otras palabras, la constante $c'' > 0$ en la expresión (3.5.17) es una función monótona creciente - no acotada - de $(1-\alpha)$. Claramente, las expresiones (3.5.13) y (3.5.17) establecen una relación evidente entre las regiones de máxima densidad final y las regiones de confianza de Fieller. Específicamente y al margen de que las constantes c'' y c_0 no son necesariamente iguales para un mismo nivel de α , es claro que la diferencia básica reside en el factor $Q(\rho)$ que aparece en (3.5.17) y desde un punto de vista

estrictamente operacional. es este factor el que evita que las regiones bayesianas presenten los inconvenientes que aparecen en las regiones de Fieller debido al hecho de que $h(\rho)$ es una función acotada de ρ . En efecto, es inmediato que si $|\rho| \rightarrow \infty$.

$$\lim G(\rho) = \infty$$

y por tanto, la magnitud

$$G(\rho) [\eta + h(\rho)]^{(\eta+r)}$$

es no acotada como función de ρ . Entonces, para todo $\alpha \in (0, 1)$ existe $M > 0$ tal que si $|\rho| > M$, se cumple que

$$G(\rho) [\eta + h(\rho)]^{(\eta+r)} > c^*$$

donde c^* es la constante positiva y finita asociada al valor de α . En otras palabras, para todo valor de $\alpha \in (0, 1)$ la región de máxima densidad final $R_{(1-\alpha)}$ es un subconjunto propio de la recta real.

CAPITULO 4. DISTRIBUCION DE REFERENCIA.

4.1 El análisis de referencia.

En ocasiones, el conocimiento inicial que posee un científico sobre un determinado parámetro puede ser considerado vago o escaso en comparación con la información que, respecto a ese parámetro, se espera obtener de un futuro experimento. Bajo tales circunstancias es común, sea cual sea la representación que se produzca del conocimiento inicial, que la distribución final resulte dominada por la verosimilitud. Esta apreciación constituye, en términos informales, una expresión del principio de estimación estable enunciado por Edwards, Lindman y Savage (1953).

De cualquier forma, tiene sentido plantearse la cuestión de si en esas condiciones merece la pena el esfuerzo, que puede ser considerable especialmente si el parámetro bajo consideración es multidimensional, de asignar una distribución que describa exactamente lo que el científico pretende saber *a priori* sobre el parámetro; o si por el contrario, resulta más razonable desde un punto de vista operativo, utilizar una distribución inicial que de acuerdo a alguna convención apropiada describa un conocimiento inicial vago.

Este argumento justifica por si solo la existencia de las llamadas distribuciones no-informativas o mínimo-informativas que si bien no forman parte del paradigma bayesiano, constituyen un auxiliar muy importante en su aplicación.

Existen además otras razones complementarias, particularmente en el ámbito de la investigación científica, que pueden sugerir el empleo de distribuciones iniciales de ese tipo. Así por ejemplo, en Box y Tiao (1973) se elabora sobre esta idea en términos similares a los siguientes:

Cuando se trata de una investigación científica, en general es razonable suponer que la verosimilitud domina a la distribución inicial puesto que, por una parte, una investigación de este tipo usualmente no se lleva a cabo si no se espera obtener información considerablemente mas precisa que la disponible previamente. Además, aun cuando el científico posea fuerte evidencia inicial sobre el parámetro de interés, resulta conveniente incluir en el reporte, junto con la distribución final que corresponde a su distribución inicial personal o subjetiva, una distribución final de referencia que corresponda a una inicial dominada por la verosimilitud ya que de esta manera es claro hasta que grado sus resultados dependen de su información inicial.

El segundo punto de esta argumentación ha sido también planteado por otros autores (Dickey 1973, por ejemplo) y en un mayor grado de generalidad, plantea que todo reporte científico debe establecer la dependencia de los resultados finales respecto a la distribución inicial considerando para tal fin un análisis de sensibilidad de la distribución final respecto a un amplio rango de posibilidades para la distribución inicial. Entre esas posibilidades se considera conveniente incluir, como referencia,

una inicial que sea dominada por la verosimilitud para así evaluar la importancia relativa de la información inicial.

En resumen, las distribuciones iniciales mínimo informativas tienen una importancia práctica considerable en la aplicación del paradigma bayesiano, especialmente a través de las distribuciones finales de referencia a que dan lugar.

Los intentos de producir distribuciones de este tipo pueden rastrearse hasta Bayes (1763) que basado en el llamado principio de razón insuficiente sugiere el empleo de iniciales uniformes. Más recientemente han aparecido otras aproximaciones a este problema entre las que destacan las distribuciones que se obtienen como límite de iniciales conjugadas (DeGroot 1970, capítulo 10), las que se producen mediante la regla de Jeffreys (Box y Tiao 1973, sección 1.3.5) y las distribuciones de referencia propuestas por Bernardo (1979). La literatura sobre el tema es abundante y una amplia recopilación bibliográfica puede encontrarse precisamente en el trabajo de Bernardo.

El concepto de distribución mínimo informativa es intuitivamente claro pero formalmente resulta fugaz y evasivo. Como ya se ha mencionado, es necesario establecer una convención respecto a la cual se pueda afirmar que una distribución describe un estado de conocimiento vago. Cada una de las aproximaciones citadas utiliza una convención distinta y aunque en ocasiones producen resultados comunes, este no es en general el caso. En este trabajo se utilizará el procedimiento debido a Bernardo (1979) que utiliza explícitamente una medida de información para comparar la evidencia disponible antes y después del experimento. Además, mientras que en los casos más simples y no conflictivos este método produce resultados que coinciden con la regla de Jeffreys y conciden o son muy similares a los de familias conjugadas, en

algunos otros casos en que tanto la regla de Jeffreys como familias conjugadas dan lugar a distribuciones finales que resultan conflictivas - porque producen resultados inadmisibles o conducen a paradojas de marginalización - las distribuciones de referencia de Bernardo producen resultados satisfactorios. Una discusión detallada del procedimiento y de estas afirmaciones puede consultarse en el artículo original de Bernardo. Aquí, simplemente se describirán las características generales del procedimiento.

La idea fundamental es la de maximizar una medida de información. Si se considera que la magnitud $I^{\circ}\{z, p(\theta)\}$ mide apropiadamente la información que se espera obtener acerca del parámetro θ cuando se realiza un experimento z y la densidad inicial de θ es $p(\theta)$ y si además, $z(s)$ denota una sucesión de s repeticiones independientes del experimento z , entonces se puede argumentar que $z(\omega)$, definido como el límite de $z(s)$ cuando $s \rightarrow \omega$, debería proveer información perfecta sobre θ y por tanto la correspondiente magnitud $I^{\circ}\{z(\omega), p(\theta)\}$, que recibe el nombre de información faltante respecto a θ , mediría la cantidad de información que se requiere para determinar con precisión el valor de θ cuando se parte de la inicial $p(\theta)$.

Intuitivamente es natural definir, dado un experimento fijo z , como mínimo informativa a la densidad $\pi(\theta)$ que maximiza la información faltante respecto a θ . El procedimiento para maximizar la funcional $I^{\circ}\{z(\omega), p(\theta)\}$ es en general complicado y requiere de un tratamiento cuidadoso; sin embargo, si se adopta una medida de información logarítmica (cf. Shannon 1948 y Lindley 1956)

$$I^{\circ}\{z(\omega), p(\theta)\} = \int p(z) \int p(\theta | z) \log[p(\theta | z)/p(\theta)] d\theta dz$$

... (4.1.1)

en donde

$$p(z) = \int p(z | \theta) p(\theta) d\theta \quad (4.1.2)$$

y z son los datos producidos por el experimento s . la idea de maximizar la información faltante ha conducido a Bernardo a proponer una densidad inicial de la forma

$$\pi(\theta) = \lim_{s \rightarrow \infty} \pi_s(\theta) \quad (4.1.3)$$

en donde

$$\pi_s(\theta) \propto \exp\left\{- \int p(z_{(s)} | \theta) Y[p(\theta | z_{(s)})] dz_{(s)}\right\} \quad \dots(4.1.4)$$

con

$$Y[g(w)] = - \int g(w) \ln[g(w)] dw \quad (4.1.5)$$

el conocido operador entropía y $z_{(s)}$ los datos producidos por $s(s)$. Más aún, si se cumplen las condiciones de regularidad necesarias para garantizar que la distribución final es asintóticamente Normal (Walker 1969, Johnson 1970 y Chen 1985 entre otros), entonces para s suficientemente grande, $p(\theta | z_{(s)})$ se puede aproximar por una Normal con media $\hat{\theta}_s$, el estimador de máxima verosimilitud de θ basado en $z_{(s)}$ y matriz de precisión $sF(\hat{\theta}_s)$ donde $F(\hat{\theta}_s)$ es la matriz de información de Fisher para θ evaluada en $\hat{\theta}_s$.

Bajo esas condiciones y utilizando un argumento límite es posible probar que (4.1.3) y (4.1.4) conducen a

$$\pi(\theta) \propto |F(\theta)|^{1/2} \quad (4.1.6)$$

de manera que el procedimiento de Bernardo coincide con la regla de Jeffreys. Esta coincidencia, sin embargo, desaparece si existen parámetros marginales. La extensión propuesta para este caso por Bernardo establece que si $\theta^t = (\theta_1^t, \theta_2^t)$ con θ_1 el parámetro de interés y la matriz de información de Fisher se escribe como

$$F(\theta) = \begin{bmatrix} F_{11}(\theta) & F_{12}(\theta) \\ F_{21}(\theta) & F_{22}(\theta) \end{bmatrix} \quad (4.1.7)$$

donde la partición de la matriz $F(\theta)$ corresponde con las dimensiones de los vectores θ_1 y θ_2 respectivamente de manera que asintóticamente,

$$\theta_1 \sim NC_{\hat{\theta}_{1n}}(F_{11}(\hat{\theta}_n)) \quad (4.1.8)$$

$$\theta_2 \sim NC_{\hat{\theta}_{2n}}(F_{22}(\hat{\theta}_n))$$

entonces, la densidad inicial de referencia $\pi(\theta_1, \theta_2)$ está dada por

$$\pi(\theta_1, \theta_2) = \pi(\theta_2 | \theta_1) \pi(\theta_1) \quad (4.1.9)$$

en donde

$$\pi(\theta_1) \propto \{|F(\theta)| / |F_{22}(\theta)|\}^{1/2} \quad (4.1.10)$$

y

$$\pi(\theta_2 | \theta_1) \propto |F_{22}(\theta)|^{1/2} \quad (4.1.11)$$

Debe observarse que puesto que en (4.1.10) el símbolo de proporcionalidad se interpreta como función de θ_1 mientras que en

(4.1.11) se interpreta como función de θ_2 , no es cierto en general que

$$\begin{aligned} \pi(\theta_1, \theta_2) &= \pi(\theta_2 | \theta_1) \pi(\theta_1) \\ &\propto |FC\theta|^{1/2} \end{aligned}$$

como establece la regla de Jeffreys para el caso multivariado. De esta forma, en este capítulo se utilizara como densidad conjunta inicial mínimo informativa o de referencia la que se obtiene de las relaciones (4.1.9), (4.1.10) y (4.1.11).

4.2 Normalidad asintótica bajo transformaciones.

El procedimiento propuesto por Bernardo (1979) permite, a partir de expresiones análogas a (4.1.4) definir, dado un experimento y un parámetro de interés, la correspondiente distribución de referencia aún cuando no se recurra a la Normalidad asintótica. Sin embargo, tal como se ha descrito en la sección anterior, esta propiedad - la Normalidad asintótica de la final - facilita enormemente la derivación de expresiones analíticas para la inicial de referencia. El estudio de las condiciones bajo las cuales una distribución final es asintóticamente Normal ha sido considerado por diversos autores (Johnson 1967 y 1970, Walker 1969, Dawid 1970, DeGroot 1970, Heyde y Johnstone 1979, Bermúdez 1985 y Chen 1985 por ejemplo) y en términos generales guardan cierta similitud con las condiciones que garantizan la Normalidad asintótica de los estimadores de máxima verosimilitud (cf. Cramér 1946, Wald 1949, LeCam 1953 y 1956 por ejemplo).

Básicamente, cada contribución en este tema establece

una serie de condiciones suficientes para asegurar la Normalidad asintótica que típicamente resultan difíciles de verificar especialmente cuando se considera una parametrización no-estándar. Más importante aún, y a diferencia de lo que ocurre con los estimadores máximo-verosímiles, no existe en la literatura, hasta donde el autor sabe, un resultado que establezca condiciones suficientes para preservar la distribución asintótica Normal bajo transformaciones - reparametrizaciones -. Es por esta causa que en esta sección se presenta un teorema que establece un conjunto de tales condiciones que aplicadas en la siguiente sección permitirán una utilización simple del método de Bernardo para obtener la distribución inicial de referencia. Primero se probarán dos lemas simples pero útiles.

Lema 1. Sea $\{X_n ; n=1, 2, \dots\}$ una sucesión de cantidades aleatorias tales que asintóticamente $X_n \sim N(\mu_n, \sigma_n^2)$ en donde las sucesiones de reales $\{\mu_n ; n=1, 2, \dots\}$ y $\{\sigma_n ; n=1, 2, \dots\}$ son tales que $\sigma_n \rightarrow 0$ cuando $n \rightarrow \infty$. Entonces, $X_n \rightarrow \mu_n$ en probabilidad.

Demostración. Dados $\delta > 0$ y $\epsilon > 0$ se tiene que

$$P[|X_n - \mu_n| < \delta] = P[|X_n - \mu_n| / \sigma_n < \delta / \sigma_n] \quad (4.2.1)$$

Sea $c > 0$ una constante tal que $\Phi(-c) < \epsilon/4$, en donde $\Phi(\cdot)$ representa, como es usual, la distribución Normal estándar $N(0, 1)$. Si ahora se selecciona $\epsilon^* < \epsilon/2$ y utilizando la Normalidad asintótica se toma n_1 tal que para todo $n > n_1$

$$|P[|X_n - \mu_n| / \sigma_n < c] - (1 - 2\Phi(-c))| < \epsilon^*. \quad (4.2.2)$$

sea n_2 tal que $\delta / \sigma_n > c$ para todo $n > n_2$ y considérese ahora un valor de $n > \max\{n_1, n_2\}$. De (4.2.1) y (4.2.2) se sigue que

$$\begin{aligned}
 P[|X_n - \mu_n| < \delta] &> P[|X_n - \mu_n| / \sigma_n < c] \\
 &> (1 - 2\Phi(-c)) - \epsilon^0 \\
 &= 1 - (2\Phi(-c) + \epsilon^0) \\
 &> 1 - \epsilon.
 \end{aligned}$$

Por lo tanto, $X_n \rightarrow \mu_n$ en probabilidad y el lema queda establecido. \square

Lema 2. Sea $\{\Sigma_n ; n=1, 2, \dots\}$ una sucesión de matrices $(k \times k)$ simétricas y positivas definidas tales que $\lambda_{n(k)} \rightarrow 0$ y $\lambda_{n(k)} = O(\lambda_{n(s)})$ cuando $n \rightarrow \infty$ en donde $\lambda_{n(k)} \geq \dots \geq \lambda_{n(s)}$ denotan los valores propios ordenados de Σ_n . Si ahora se consideran las sucesiones $\{\underline{a}_n ; n=1, 2, \dots\}$ y $\{\underline{b}_n ; n=1, 2, \dots\}$ de vectores en \mathbb{R}^k tales que $\underline{a}_n - \underline{b}_n \rightarrow \underline{0}$ cuando $n \rightarrow \infty$ y $\|\underline{b}_n\| > m$ para alguna constante positiva m si n es suficientemente grande, entonces

$$\lim_{n \rightarrow \infty} \frac{(\underline{a}_n^t \Sigma_n \underline{a}_n)}{(\underline{b}_n^t \Sigma_n \underline{b}_n)} = 1.$$

Demostración. Dado que Σ_n es, por hipótesis, simétrica es posible escribir

$$\Sigma_n = P_n D_n P_n^t$$

donde

$$D_n = \text{diag}\{\lambda_{n(s)}, \dots, \lambda_{n(k)}\}$$

y P_n es tal que

$$P_n^t P_n = P_n P_n^t = I.$$

Entonces.

$$\begin{aligned}
 \langle \mathbf{a}_n^t \Sigma_n \mathbf{a}_n \rangle / \langle \mathbf{b}_n^t \Sigma_n \mathbf{b}_n \rangle &= \langle \mathbf{a}_n^t \mathbf{P}_n \mathbf{D}_n \mathbf{P}_n^t \mathbf{a}_n \rangle / \langle \mathbf{b}_n^t \mathbf{P}_n \mathbf{D}_n \mathbf{P}_n^t \mathbf{b}_n \rangle \\
 &= \langle \mathbf{P}_n^t \mathbf{a}_n \rangle^t \mathbf{D}_n \langle \mathbf{P}_n^t \mathbf{a}_n \rangle / \langle \mathbf{P}_n^t \mathbf{b}_n \rangle^t \mathbf{D}_n \langle \mathbf{P}_n^t \mathbf{b}_n \rangle \\
 &= \langle \mathbf{y}_n^t \mathbf{D}_n \mathbf{y}_n \rangle / \langle \mathbf{u}_n^t \mathbf{D}_n \mathbf{u}_n \rangle
 \end{aligned}$$

... (4.2.3)

con $\mathbf{y}_n = \mathbf{P}_n^t \mathbf{a}_n$ y $\mathbf{u}_n = \mathbf{P}_n^t \mathbf{b}_n$. Es claro que

$$\|\mathbf{u}_n\| = \|\mathbf{b}_n\|$$

$$\|\mathbf{y}_n\| = \|\mathbf{a}_n\|$$

$$\|\mathbf{y}_n - \mathbf{u}_n\| = \|\mathbf{a}_n - \mathbf{b}_n\|$$

de manera que $\mathbf{y}_n - \mathbf{u}_n \rightarrow 0$ cuando $n \rightarrow \infty$. Más aún, utilizando el hecho de que $\|\mathbf{u}_n\| > m > 0$ si n es suficientemente grande, se tiene que

$$\begin{aligned}
 \left| \frac{\|\mathbf{y}_n\|}{\|\mathbf{u}_n\|} - 1 \right| &= \left| \frac{\|\mathbf{y}_n\| - \|\mathbf{u}_n\|}{\|\mathbf{u}_n\|} \right| \\
 &= \frac{|\|\mathbf{y}_n\| - \|\mathbf{u}_n\||}{\|\mathbf{u}_n\|} \\
 &\leq \frac{\|\mathbf{y}_n - \mathbf{u}_n\|}{\|\mathbf{u}_n\|} \\
 &\leq \frac{\|\mathbf{y}_n - \mathbf{u}_n\|}{m}
 \end{aligned}$$

Por lo tanto, se tiene que

$$\lim_{n \rightarrow \infty} \frac{\|v_n\|}{\|u_n\|} = 1.$$

Ahora bien.

$$\begin{aligned} |(a_n^t \Sigma_{n,n}^{-1}) / (b_n^t \Sigma_{n,n}^{-1}) - 1| &= | \langle v_n^t D_n v_n - u_n^t D_n u_n \rangle / \langle u_n^t D_n u_n \rangle | \\ &= | (\sum_{l=1}^k \lambda_{nl} v_{nl}^2 - \sum_{l=1}^k \lambda_{nl} u_{nl}^2) / (\sum_{l=1}^k \lambda_{nl} u_{nl}^2) | \\ &= | (\sum_{l=1}^k \lambda_{nl} (v_{nl}^2 - u_{nl}^2)) / (\sum_{l=1}^k \lambda_{nl} u_{nl}^2) | \\ &= | \sum_{l=1}^k \lambda_{nl} (v_{nl}^2 - u_{nl}^2) | / \sum_{l=1}^k \lambda_{nl} u_{nl}^2 \\ &\dots (4.2.4) \end{aligned}$$

Por lo tanto,

$$|(a_n^t \Sigma_{n,n}^{-1}) / (b_n^t \Sigma_{n,n}^{-1}) - 1| \leq (\lambda_{n(k)} / \lambda_{n(1)}) \sum_{l=1}^k |v_{nl}^2 - u_{nl}^2| / \|u_n\|^2$$

... (4.2.5)

pero por hipótesis $\lambda_{n(k)} = O(\lambda_{n(1)})$, de manera que existe una constante $M > 0$ tal que $\lambda_{n(k)} \leq M \lambda_{n(1)}$ si n es suficientemente grande. Así, para n suficientemente grande se tiene que

$$(\lambda_{n(k)} / \lambda_{n(1)}) \sum_{l=1}^k |v_{nl}^2 - u_{nl}^2| / \|u_n\|^2 \leq M \sum_{l=1}^k |v_{nl}^2 - u_{nl}^2| / \|u_n\|^2$$

... (4.2.6)

Por otra parte,

$$|v_{ni}^2 - u_{ni}^2| / \|u_n\|^2 = (|v_{ni} - u_{ni}| / \|u_n\|) (|v_{ni} + u_{ni}| / \|u_n\|) \dots (4.2.7)$$

en donde

$$\begin{aligned} |v_{ni} + u_{ni}| / \|u_n\| &\leq (|v_{ni}| + |u_{ni}|) / \|u_n\| \\ &\leq (\|v_n\| + \|u_n\|) / \|u_n\| \\ &= 1 + \|v_n\| / \|u_n\| \dots (4.2.8) \end{aligned}$$

En consecuencia, si n es suficientemente grande, de las expresiones (4.2.5), (4.2.6), (4.2.7) y (4.2.8) se sigue que

$$| \langle a_n^t \Sigma_n a_n \rangle / \langle b_n^t \Sigma_n b_n \rangle - 1 | \leq (M/m) (1 + \|v_n\| / \|u_n\|) \sum_{i=1}^k |v_{ni} - u_{ni}|$$

de donde, puesto que $\|v_n\| / \|u_n\| \rightarrow 0$ y $|v_{ni} - u_{ni}| \rightarrow 0$ si $n \rightarrow \infty$, se obtiene la conclusión del lema. \square

Contando con estos resultados preliminares es posible proceder a establecer un primer

Teorema 1. Sea $\{X_n; n=1,2,\dots\}$ una sucesión de vectores aleatorios en \mathbb{R}^k . Suponga que asintóticamente $X_n \sim N(\underline{\mu}_n, \Sigma_n)$ con $\{\underline{\mu}_n; n=1,2,\dots\}$ una sucesión de vectores en \mathbb{R}^k tales que para un $\underline{\mu} \in \mathbb{R}^k$ fijo, $\underline{\mu}_n \rightarrow \underline{\mu}$, $\{\Sigma_n; n=1,2,\dots\}$ una sucesión de matrices de covarianzas $(k \times k)$ simétricas y positivas definidas tales que $\lambda_{n(k)} \rightarrow 0$ y $\lambda_{n(1)} = O(\lambda_{n(k)})$ si $n \rightarrow \infty$ en donde $\lambda_{n(k)}$ y $\lambda_{n(1)}$ denotan respectivamente el máximo y el mínimo de los valores propios de Σ_n . Si $\Psi(X_n) = (\psi_1(X_n), \dots, \psi_r(X_n))^t$ ($r \leq k$) es una función tal que la matriz

$$J(\underline{t}) = \left[\frac{\partial \psi_i}{\partial X_j} \right]_{\underline{x}=\underline{t}} \quad (r \times k)$$

es no singular en $\underline{t} = \underline{\mu}$ y tiene entradas continuas también en $\underline{t} = \underline{\mu}$. entonces asintóticamente

$$\underline{\psi}(X_n) \sim NC(\underline{\psi}_n, J_n \Sigma_n J_n^t)$$

en donde $\underline{\psi}_n = \underline{\psi}(\underline{\mu}_n)$ y $J_n = J(\underline{\mu}_n)$.

Demostración. Puesto que por definición una sucesión de vectores aleatorios $\{Y_n\}$ asintóticamente tiene una distribución Normal multivariada $NC(\underline{\mu}_n, \Sigma_n)$ si para todo vector \underline{d} tal que $\underline{d}^t \Sigma_n \underline{d} \geq 0$ para toda n suficientemente grande, la sucesión $\{\underline{d}^t Y_n\}$ es asintóticamente Normal univariada $NC(\underline{d}^t \underline{\mu}_n, \underline{d}^t \Sigma_n \underline{d})$, y dado que para todo vector \underline{d} , $\underline{d}^t Y_n = \underline{d}^t \underline{\psi}(X_n)$ es una función real de X_n , es suficiente con probar la versión más simple del teorema ($r=1$) sustituyendo la matriz $J(\underline{t})$ por el vector de derivadas parciales

$$g(\underline{t}) = \left[\frac{\partial \psi}{\partial X_j} \right]_{\underline{x}=\underline{t}} \quad (1 \times k)$$

y la no singularidad de $J(\underline{t})$ por la condición $g_0 = g(\underline{\mu}) \neq 0$. En esta forma, aplicando el Teorema de Taylor alrededor de $\underline{\mu}_n$ y escribiendo $Y_n = \underline{\psi}(X_n)$, $\underline{\psi}_n = \underline{\psi}(\underline{\mu}_n)$ se tiene que

$$Y_n = \underline{\psi}(\underline{\mu}_n) + g(\underline{\xi}_n) (X_n - \underline{\mu}_n) \quad (4.2.9)$$

en donde $\underline{\xi}_n = \gamma_n X_n + (1 - \gamma_n) \underline{\mu}_n$ para algún $\gamma_n \in (0, 1)$. Si n es suficientemente grande, la continuidad de g garantiza que $g_n \Sigma_n g_n^t > 0$ de manera que se puede escribir

$$Z_n = [Y_n - \underline{\psi}(\underline{\mu}_n)] / [g_n \Sigma_n g_n^t]^{1/2}$$

$$= [g(\xi_n)(x_n - \mu_n)] / [e_n \Sigma_n e_n^t]^{1/2}$$

$$= \{ [e_0 \Sigma_n e_0^t]^{1/2} / [e_n \Sigma_n e_n^t]^{1/2} H[e_n^*(x_n - \mu_n)] / [e_0 \Sigma_n e_0^t]^{1/2}$$

con $e_n^* = g(\xi_n)$. Puesto que $\lim \mu_n = \mu$ y g es continua, se sigue que $\lim e_n^* = e_0^*$ o equivalentemente, $\lim (e_n^* - e_0^*) = 0$. Además $e_0^* \neq 0$, así que una aplicación del Lema 2 produce el siguiente resultado

$$\lim_{n \rightarrow \infty} \{ [e_0 \Sigma_n e_0^t]^{1/2} / [e_n \Sigma_n e_n^t]^{1/2} \} = 1.$$

Por lo tanto utilizando el Teorema de Slutsky (Serfling 1980, sección 1.5.4), se tiene que la distribución asintótica de Z_n coincide con la de la sucesión de cantidades aleatorias

$$W_n = [e_n^*(x_n - \mu_n)] / [e_0 \Sigma_n e_0^t]^{1/2}$$

si tal distribución límite existe. Para obtener la distribución asintótica de W_n es conveniente escribir

$$W_n = [e_0^*(x_n - \mu_n)] / [e_0 \Sigma_n e_0^t]^{1/2} + [(e_n^* - e_0^*)(x_n - \mu_n)] / [e_0 \Sigma_n e_0^t]^{1/2}$$

$$= \alpha_n + \beta_n \quad (4.2.10)$$

El primer término en el lado derecho de (4.2.10) es, por hipótesis, asintóticamente distribuida como $N(0, 1)$. Por lo tanto, la demostración del teorema puede concluirse probando que el segundo término en el lado derecho de (4.2.10) es insignificante en comparación con α_n . Más específicamente, se probará que β_n converge a cero en probabilidad. Primero, se comprobará que $\xi_n \rightarrow \mu$ en probabilidad. Directamente de la definición de ξ_n se tiene que

$$\begin{aligned}
 \|\underline{x}_n - \underline{\mu}\| &= \|\underline{x}_n - \underline{\mu}_n + \underline{\mu}_n - \underline{\mu}\| \\
 &\leq \|\underline{x}_n - \underline{\mu}_n\| + \|\underline{\mu}_n - \underline{\mu}\| \\
 &= \|\gamma_n \underline{x}_n + (1 - \gamma_n) \underline{\mu}_n - \underline{\mu}_n\| + \|\underline{\mu}_n - \underline{\mu}\| \\
 &= \|\gamma_n (\underline{x}_n - \underline{\mu}_n)\| + \|\underline{\mu}_n - \underline{\mu}\| \\
 &= |\gamma_n| \|\underline{x}_n - \underline{\mu}_n\| + \|\underline{\mu}_n - \underline{\mu}\| \\
 &\leq \|\underline{x}_n - \underline{\mu}_n\| + \|\underline{\mu}_n - \underline{\mu}\|
 \end{aligned}$$

así que una aplicación repetida del Lema 1 a las entradas $\langle X_{ni} - \mu_{ni} \rangle$; $i = 1, \dots, k$ de $\langle \underline{x}_n - \underline{\mu}_n \rangle$ y el hecho de que $\underline{\mu}_n \rightarrow \underline{\mu}$ determinísticamente, completan la prueba. Ahora, se tiene que

$$\begin{aligned}
 |w_n| &= |[\langle \underline{e}_n^0 - \underline{e}_0 \rangle \langle \underline{x}_n - \underline{\mu}_n \rangle] / [\langle \underline{e}_0 \Sigma_n \underline{e}_0^t \rangle]^{1/2}| \\
 &= |[\langle \underline{e}_n^0 - \underline{e}_0 \rangle \langle \underline{x}_n - \underline{\mu}_n \rangle]| / [\langle \underline{e}_0 \Sigma_n \underline{e}_0^t \rangle]^{1/2} \\
 &= |[\sum_{i=1}^k \langle \rho_{ni}^0 - \rho_{0i} \rangle \langle X_{ni} - \mu_{ni} \rangle]| / [\langle \underline{e}_0 \Sigma_n \underline{e}_0^t \rangle]^{1/2} \\
 &\leq [\sum_{i=1}^k |\rho_{ni}^0 - \rho_{0i}| |X_{ni} - \mu_{ni}|] / [\langle \underline{e}_0 \Sigma_n \underline{e}_0^t \rangle]^{1/2} \\
 &= \sum_{i=1}^k |\rho_{ni}^0 - \rho_{0i}| [|X_{ni} - \mu_{ni}| / [\langle \underline{e}_0 \Sigma_n \underline{e}_0^t \rangle]^{1/2}]
 \end{aligned}$$

Si entonces para cada $i=1, \dots, k$, σ_{ni}^2 denota el i -ésimo elemento en la diagonal de Σ_n y \underline{e}_i denota, como ya es usual, el i -ésimo vector de la base canónica en \mathbb{R}^k , se tiene que la cantidad aleatoria

$$S_{ni} = (X_{ni} - \mu_{ni}) / \sigma_{ni}$$

es, por hipótesis, asintóticamente distribuida como NCO. 1) y por tanto $|S_{ni}|$ tiene a su vez una distribución límite bien definida.

Por otra parte,

$$\begin{aligned} \sigma_{ni} / [e_0 \Sigma_n e_0^t]^{1/2} &= [(\sigma_{ni})^2 / [e_0 \Sigma_n e_0^t]]^{1/2} \\ &= \{ [e_i \Sigma_n e_i] / [e_0 \Sigma_n e_0^t] \}^{1/2} \end{aligned}$$

... (4.2.11)

y puesto que

$$\begin{aligned} \frac{e_i \Sigma_n e_i}{n} &\leq \lambda_{n(k)} \|e_i\|^2 \\ &= \lambda_{n(k)} \end{aligned}$$

y también

$$e_0 \Sigma_n e_0^t \geq \lambda_{n(s)} \|e_0\|^2$$

se sigue que

$$\sigma_{ni} / [e_0 \Sigma_n e_0^t]^{1/2} \leq [\lambda_{n(k)} / \lambda_{n(s)}]^{1/2} / \|e_0\|$$

Por lo tanto y como $\lambda_{n(k)} = O(\lambda_{n(k)})$ si $n \rightarrow \infty$, se sigue que la magnitud (4.2.11) está acotada. Finalmente, puesto que $|\varphi_{ni}^* - \varphi_{oi}|$ converge a cero, una extensión del Teorema de Slutsky garantiza que

$$|\beta_{ni}| = |\varphi_{ni}^* - \varphi_{oi}| |X_{ni} - \mu_{ni}| / [\varrho_0 \Sigma_n \varrho_0^t]^{1/2}$$

converge a cero en probabilidad. En consecuencia, $\beta_n \rightarrow 0$ en probabilidad y el teorema queda establecido. \square

Observación 1. En el caso $k=1$, esto es cuando se considera una función real de una magnitud real, la condición sobre los valores propios de Σ_n es obviamente correcta y el resultado es, en ese sentido, más fuerte.

Observación 2. El Teorema 1 es una generalización de algunos resultados muy conocidos en la literatura (Serfling 1980, sección 3.3) en donde $\mu_n = \underline{\mu}$ para toda n y $\Sigma_n = b_n^2 \Sigma$ con $\{b_n\}$ una sucesión de reales tales que $b_n \rightarrow 0$ y Σ es una matriz de covarianzas fija. Este resultado sólo requiere la hipótesis adicional de continuidad de las derivadas parciales de la transformación en el punto $\underline{t} = \underline{\mu}$.

En un contexto bayesiano, la Normalidad asintótica está típicamente referida a una distribución final para un parámetro dado y la situación es ligeramente distinta de la que cubre el Teorema 1. El tipo de convergencia considerado en este teorema es, en cierto sentido, una convergencia *determinística* de medidas de probabilidad mientras que el tipo de resultado que se emplea para una distribución final es una convergencia *probabilística* - con respecto al espacio muestral - de medidas de probabilidad definidas sobre el espacio de parámetros. Esta situación ha sido reconocida en la literatura y de hecho, se ha utilizado para

facilitar la prueba de algunos resultados (cf. Chen 1985, por ejemplo).

La reformulación del Teorema 1, con la terminología probabilística necesaria para cubrir la aplicación en el ámbito bayesiano es inmediata.

Teorema 2. Sea $\underline{\theta} \in \Theta$ ($\Theta \subseteq \mathbb{R}^k$) un vector de parámetros. Sea $\pi_n(\underline{\theta}) = \pi(\underline{\theta} | \underline{X}_{(n)})$ la distribución final de $\underline{\theta}$ después de haber observado la muestra $\underline{X}_{(n)} = \{X_1, \dots, X_n\}$ del modelo $p(X | \underline{\theta}_0)$. Suponga que la sucesión $\{\pi_n(\underline{\theta})\}$ es tal que con probabilidad 1 respecto a $p(\underline{X}_{(n)} | \underline{\theta}_0)$ y condicionado a los valores observados en la muestra $\underline{X}_{(n)}$, $\underline{\theta}$ tiene una distribución asintótica Normal $N(\hat{\underline{\theta}}_n(\underline{X}_{(n)}), \Sigma_n(\underline{X}_{(n)}))$ en donde $\hat{\underline{\theta}}_n(\underline{X}_{(n)}) \rightarrow \underline{\theta}_0$ en probabilidad respecto a $p(\underline{X}_{(n)} | \underline{\theta}_0)$, $\Sigma_n(\underline{X}_{(n)})$ es una matriz de covarianzas simétrica y positiva definida para toda n cuyos valores propios máximo ($\lambda_{n(k)}$) y mínimo ($\lambda_{n(1)}$) satisfacen las propiedades $\lambda_{n(k)} \rightarrow 0$ y $\lambda_{n(k)} = O(\lambda_{n(1)})$ si $n \rightarrow \infty$; ambas condiciones en probabilidad con respecto a la medida $p(\underline{X}_{(n)} | \underline{\theta}_0)$. Si $\underline{\psi}(\underline{\theta}) = (\psi_1(\underline{\theta}), \dots, \psi_r(\underline{\theta}))^t$ ($r \leq k$) es tal que la matriz de derivadas parciales

$$J(\underline{t}) = \left[\frac{\partial \psi_i}{\partial \theta_j} \right]_{\underline{t}=\underline{\theta}}$$

es no singular en el punto $\underline{t} = \underline{\theta}_0$ y tiene entradas continuas también en $\underline{t} = \underline{\theta}_0$, entonces con probabilidad 1 respecto a la medida $p(\underline{X}_{(n)} | \underline{\theta}_0)$, la distribución de $\underline{\psi}(\underline{\theta})$ dada $\underline{X}_{(n)}$ resulta asintóticamente Normal $N(\underline{\psi}(\hat{\underline{\theta}}_n), J_n^{-1} \Sigma_n(\underline{X}_{(n)}) J_n)$ con $J_n = J(\hat{\underline{\theta}}_n)$.

Demostración. Es una consecuencia del Teorema 1.

Observación 3. Debe notarse que las condiciones que se imponen sobre la transformación son las mismas en ambos teoremas.

Las diferencias aparecen en el comportamiento de las sucesiones $\hat{\theta}_n$ y Σ_n . En el Teorema 2, $\hat{\theta}_n$ es un estimador consistente de θ_0 y Σ_n tiende a la matriz cero si $n \rightarrow \infty$, también en probabilidad. Este tipo de condiciones son usualmente verificadas cuando se comprueba la Normalidad asintótica para una parametrización particular (cf. Walker 1969, Dawid 1970, Fraser y McDunnough 1984, y Chen 1985). La única condición nueva en ambos teoremas es la caracterización de la velocidad relativa de convergencia de los valores propios de Σ_n . Es necesario reconocer que esta condición puede ser muy difícil de comprobar, especialmente si la dependencia funcional de Σ_n respecto $X_{(n)}$ es complicada. Afortunadamente, a menudo ocurre que $\Sigma_n(X_{(n)}) = \Sigma_n(\hat{\theta}_n)$ y aún más, la matriz $[\Sigma_n(\hat{\theta}_n)]^{-1}$ usualmente coincide con la matriz de información de Fisher para θ , de manera que para un conjunto de observaciones independientes e idénticamente distribuidas $X_{(n)}$, se tiene que $\Sigma_n(\hat{\theta}_n) = [\Sigma(\hat{\theta}_n)] / n$ en donde $\Sigma(\theta)$ no varía con n . Así, la estructura de los valores propios de $\Sigma_n(X_{(n)})$ es la misma que la de $\Sigma(\hat{\theta}_n)$. Finalmente, si $\Sigma(\theta)$ es una función continua de θ , como típicamente es el caso, se tiene que $\Sigma(\hat{\theta}_n) \rightarrow \Sigma(\theta_0)$ y entonces, la continuidad de la estructura de los valores propios (Anderson, Bröns y Jensen 1983, Lema 2) reduce el problema a la consideración de la matriz $\Sigma(\theta_0)$ para verificar que es de rango completo, un problema de solución inmediata.

Como comentario final, es conveniente insistir en que las condiciones que ambos teoremas imponen son suficientes pero es posible que no sean necesarias. Un estudio más detallado sería necesario en esa dirección.

4.3 Distribución inicial de referencia para el cociente de combinaciones lineales.

Una vez establecida la naturaleza de las distribuciones de referencia y contando con los resultados previos relativos a la Normalidad asintótica bajo transformaciones, es posible obtener la distribución de referencia para el modelo que se considera en esta tesis. Recordando el modelo reparametrizado

$$\underline{Y} \sim NC(\underline{Z}\beta, \sigma^2 I) \quad (4.3.1)$$

es posible probar, utilizando por ejemplo las condiciones propuestas por DeGroot (1970, sección 10.11), que la distribución final del vector de parámetros $(\beta^t, \sigma^2)^t$, cuando se han realizado n repeticiones independientes del experimento es, con probabilidad 1, asintóticamente Normal con vector de medias $(\hat{\beta}_n^t, \hat{\sigma}_n^2)^t$ y matriz de precisión $n I(\hat{\beta}_n, \hat{\sigma}_n)$ en donde $\hat{\beta}_n$ y $\hat{\sigma}_n$ son los correspondientes estimadores de máxima verosimilitud de β y σ basados en las n repeticiones del experimento mientras que $I(\hat{\beta}_n, \hat{\sigma}_n)$ es la matriz de información de Fisher para (β, σ) evaluada en $(\hat{\beta}_n, \hat{\sigma}_n)$. Puesto que

$$L(\beta, \sigma) = (2n\sigma^2)^{-N/2} \exp\{-(\underline{Y} - \underline{Z}\beta)^t(\underline{Y} - \underline{Z}\beta)/(2\sigma^2)\}$$

se tiene que si $l(\beta, \sigma) = \log[L(\beta, \sigma)]$ entonces

$$l(\beta, \sigma) = -(N/2) \log(2n\sigma^2) - (\underline{Y} - \underline{Z}\beta)^t(\underline{Y} - \underline{Z}\beta)/(2\sigma^2)$$

de donde

$$\frac{\partial l(\beta, \sigma)}{\partial \beta} = - \{ \underline{Z}^t \underline{Z} \beta - \underline{Z}^t \underline{Y} \} / \sigma^2$$

$$\frac{\partial l(\beta, \sigma)}{\partial \sigma} = -N/\sigma + (\bar{Y} - Z\beta)^2 (\bar{Y} - Z\beta) / \sigma^3$$

y en consecuencia

$$\frac{\partial^2 l(\beta, \sigma)}{\partial \beta \partial \beta} = - (Z^T Z) \sigma^{-2}$$

$$\frac{\partial^2 l(\beta, \sigma)}{\partial \sigma \partial \sigma} = N/\sigma^2 - 3 (\bar{Y} - Z\beta)^2 (\bar{Y} - Z\beta) / \sigma^4$$

$$\begin{aligned} \frac{\partial^2 l(\beta, \sigma)}{\partial \beta \partial \sigma} &= \frac{\partial^2 l(\beta, \sigma)}{\partial \sigma \partial \beta} \\ &= -2 Z^T \{ \bar{Y} - Z\beta \} / \sigma^3 \end{aligned}$$

y más aún,

$$- E \left\{ \frac{\partial^2 l(\beta, \sigma)}{\partial \beta \partial \beta} \right\} = (Z^T Z) \sigma^{-2}$$

$$\begin{aligned} - E \left\{ \frac{\partial^2 l(\beta, \sigma)}{\partial \sigma \partial \sigma} \right\} &= - N/\sigma^2 + 3N \sigma^2 / \sigma^4 \\ &= 2N \sigma^{-2} \end{aligned}$$

$$\begin{aligned} - E \left\{ \frac{\partial^2 l(\beta, \sigma)}{\partial \beta \partial \sigma} \right\} &= - E \left\{ \frac{\partial^2 l(\beta, \sigma)}{\partial \sigma \partial \beta} \right\} \\ &= 0. \end{aligned}$$

En otras palabras, la matriz de información de Fisher está dada por

$$I(\beta, \sigma) = \sigma^{-2} \begin{bmatrix} CZ^tZ & 0 \\ 0^t & 2N \end{bmatrix} \quad (4.3.2)$$

de manera que asintóticamente, la matriz de precisión del vector de parámetros $\zeta^t = (\beta^t, \sigma)$ está dada por

$$nI(\hat{\zeta}) = nI(\hat{\beta}_n, \hat{\sigma}_n) = n \hat{\sigma}_n^{-2} \begin{bmatrix} CZ^tZ & 0 \\ 0^t & 2N \end{bmatrix} \quad (4.3.3)$$

Si ahora se considera la función que transforma al vector ζ en $(\gamma^t)^t = (\gamma_1, \gamma_2, \dots, \gamma_{k+1})$ con

$$\begin{aligned} \gamma_1 &= \zeta_1 / \zeta_2 \\ &= \beta_1 / \beta_2 \\ \gamma_i &= \zeta_i \quad ; \quad i = 1, \dots, k+1 \end{aligned}$$

se tiene que la matriz de derivadas parciales

$$J = \begin{bmatrix} \frac{\partial \gamma_j}{\partial \zeta_i} \end{bmatrix} = \begin{bmatrix} \frac{\partial \gamma_1}{\partial \zeta_1} & \dots & \frac{\partial \gamma_{k+1}}{\partial \zeta_1} \\ \frac{\partial \gamma_1}{\partial \zeta_{k+1}} & \dots & \frac{\partial \gamma_{k+1}}{\partial \zeta_{k+1}} \end{bmatrix}$$

está simplemente dada por

$$J = \begin{bmatrix} \zeta_2^{-1} & 0 & \dots & 0 \\ -\zeta_1 \zeta_2^{-2} & 1 & \dots & 0 \\ 0 & & & \\ \vdots & & & \\ 0 & \dots & \dots & 1 \end{bmatrix}$$

que está bien definida y es nosingular si $\zeta_2 \neq 0$ pero esta restricción se satisface por hipótesis. En consecuencia y puesto que se cumplen las condiciones del Teorema 2 de la sección 4.2 se tiene que, asintóticamente $\hat{\zeta}^*$ es Normal con vector de medias ζ^* y matriz de covarianzas

$$n^{-1} J_n [I_n(\hat{\theta}_n, \hat{\sigma}_n^2)]^{-1} J_n^t$$

o equivalentemente, con matriz de precisión

$$n I_n(\hat{\zeta}_n^*) = n T_n^t [I_n(\hat{\theta}_n, \hat{\sigma}_n^2)] T_n$$

en donde $T_n = J_n^{-1}$. Es inmediato comprobar que

$$T = \begin{bmatrix} \zeta_2 & \zeta_1 \zeta_2^{-1} & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & & & 1 \\ \vdots & & & \\ 0 & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_2 & \gamma_1 & \dots & 0 \\ 0 & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & & & I_{(k-1)} \end{bmatrix}$$

$$= \sigma^{-2} \begin{bmatrix} T_{11}^t (Z^t Z) T_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 2N \end{bmatrix} \quad (4.3.6)$$

Ahora bien, si se recurre a (3.4.6) se tiene que

$$Z^t Z = \begin{bmatrix} Z_1^t Z_1 & \dots & Z_1^t Z_2 \\ \vdots & \ddots & \vdots \\ Z_2^t Z_1 & \dots & Z_2^t Z_2 \end{bmatrix}$$

de manera que

$$\begin{aligned} (Z^t Z) T_{11} &= \begin{bmatrix} Z_1^t Z_1 & \dots & Z_1^t Z_2 \\ \vdots & \ddots & \vdots \\ Z_2^t Z_1 & \dots & Z_2^t Z_2 \end{bmatrix} \begin{bmatrix} A_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & I_{(k-2)} \end{bmatrix} \\ &= \begin{bmatrix} (Z_1^t Z_1) A_{11} & \dots & Z_1^t Z_2 \\ \vdots & \ddots & \vdots \\ (Z_2^t Z_1) A_{11} & \dots & Z_2^t Z_2 \end{bmatrix} \end{aligned}$$

y entonces

$$\begin{aligned} T_{11}^t (Z^t Z) T_{11} &= \begin{bmatrix} A_{11}^t & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & I_{(k-2)} \end{bmatrix} \begin{bmatrix} (Z_1^t Z_1) A_{11} & \dots & Z_1^t Z_2 \\ \vdots & \ddots & \vdots \\ (Z_2^t Z_1) A_{11} & \dots & Z_2^t Z_2 \end{bmatrix} \\ &= \begin{bmatrix} A_{11}^t (Z_1^t Z_1) A_{11} & \dots & A_{11}^t (Z_1^t Z_2) \\ \vdots & \ddots & \vdots \\ (Z_2^t Z_1) A_{11} & \dots & Z_2^t Z_2 \end{bmatrix} \end{aligned}$$

Más aún, puesto que $Z_1 = (Z_{11}, Z_{12})$ ($N \times 2$), se tiene que

$$Z_1^t Z_1 = \begin{bmatrix} N_1^t N_1 & \cdot & N_1^t N_2 \\ \cdot & \cdot & \cdot \\ N_2^t N_1 & \cdot & N_2^t N_2 \end{bmatrix}$$

de donde

$$\begin{aligned} (Z_1^t Z_1) A_{11} &= \begin{bmatrix} N_1^t N_1 & \cdot & N_1^t N_2 \\ \cdot & \cdot & \cdot \\ N_2^t N_1 & \cdot & N_2^t N_2 \end{bmatrix} \begin{bmatrix} \gamma_2 & \gamma_1 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} (N_1^t N_1) \gamma_2 & \cdot & (N_1^t N_1) \gamma_1 + (N_1^t N_2) \\ \cdot & \cdot & \cdot \\ (N_2^t N_1) \gamma_2 & \cdot & (N_2^t N_1) \gamma_1 + (N_2^t N_2) \end{bmatrix} \end{aligned}$$

y entonces

$$\begin{aligned} A_{11}^t (Z_1^t Z_1) A_{11} &= \begin{bmatrix} \gamma_2 & 0 \\ \gamma_1 & 1 \end{bmatrix} \begin{bmatrix} (N_1^t N_1) \gamma_2 & \cdot & (N_1^t N_1) \gamma_1 + (N_1^t N_2) \\ \cdot & \cdot & \cdot \\ (N_2^t N_1) \gamma_2 & \cdot & (N_2^t N_1) \gamma_1 + (N_2^t N_2) \end{bmatrix} \\ &= \begin{bmatrix} \cdot & S_{11} \gamma_2^2 & \cdot & \cdot & S_{11} \gamma_1 \gamma_2 + S_{12} \gamma_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ S_{11} \gamma_2 \gamma_1 + S_{21} \gamma_2 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & S_{11} \gamma_1^2 + 2 S_{21} \gamma_1 + S_{22} \end{bmatrix} \end{aligned}$$

con $S_{ij} = (N_i^t N_j)$; $i, j = 1, 2$. Por otra parte,

$$\begin{aligned} Z_1^t Z_2 &= \begin{bmatrix} N_1^t \\ N_2^t \end{bmatrix} Z_2 \\ &= \begin{bmatrix} N_1^t Z_2 \\ N_2^t Z_2 \end{bmatrix} \end{aligned}$$

de manera que

$$\begin{aligned}
 A_{11}^{-1}(Z_1^t Z_2) &= \begin{bmatrix} \gamma_2 & 0 \\ \gamma_1 & 1 \end{bmatrix} \begin{bmatrix} Z_1^t Z_2 \\ Z_2^t Z_2 \end{bmatrix} \\
 &= \begin{bmatrix} \gamma_2 (Z_1^t Z_2) \\ \gamma_1 (Z_1^t Z_2) + (Z_2^t Z_2) \end{bmatrix}
 \end{aligned}$$

y en consecuencia, la matriz $((k+1) \times (k+1))$ $I(\chi^*) = T^t I(\beta, \sigma) T$ está dada por la expresión

$$\sigma^2 I(\chi^*) = \begin{bmatrix} S_{11} \gamma_2^2 & S_{11} \gamma_2 \gamma_1 + S_{12} \gamma_2 & \dots & Z_1^t Z_2 \gamma_2 & \dots & 0 \\ S_{11} \gamma_1 \gamma_2 + S_{21} \gamma_2 & S_{11} \gamma_1^2 + 2S_{12} \gamma_1 + S_{22} & \dots & Z_1^t Z_2 \gamma_1 + Z_2^t Z_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Z_2^t Z_1 \gamma_1 & Z_2^t Z_1 \gamma_1 + Z_2^t Z_2 & \dots & Z_2^t Z_2 & \dots & \dots \\ \dots & \dots & \dots & 0^t & \dots & 2N \end{bmatrix}$$

Ahora bien, de (4.3.4), (4.3.5) y (4.3.6) se sigue que

$$\begin{aligned}
 |I(\chi^*)| &= \sigma^{-2(k+1)} \begin{vmatrix} T_{11}^t (Z^t Z) T_{11} & \dots & 0 \\ \dots & \dots & \dots \\ 0^t & \dots & 2N \end{vmatrix} \\
 &= 2N \sigma^{-2(k+1)} |T_{11}^t (Z^t Z) T_{11}| \\
 &= 2N \sigma^{-2(k+1)} |T_{11}^t| |Z^t Z| |T_{11}| \\
 &= 2N \sigma^{-2(k+1)} |Z^t Z| |T_{11}|^2 \\
 &= 2N \sigma^{-2(k+1)} |Z^t Z| |A_{11}|^2 \\
 &= 2N \sigma^{-2(k+1)} |Z^t Z| \gamma_2^2
 \end{aligned} \tag{4.3.7}$$

Sea ahora H la matriz $(k \times k)$ definida por la relación

y la definición de $Q(\rho)$ en (3.3.15).

$$\begin{aligned}
 |H| &= 2N\sigma^{-2k} |D|^{-1} \{ [\underline{z}_1^t(I-P)\underline{z}_1] \gamma_1^2 + 2 [\underline{z}_1^t(I-P)\underline{z}_2] \gamma_1 \\
 &\quad + [\underline{z}_2^t(I-P)\underline{z}_2] \} \\
 &= 2N\sigma^{-2k} |Z_2^t Z_2| \{ c_2 \gamma_1^2 + c_1 \gamma_1 + c_0 \} \\
 &= 2N\sigma^{-2k} |Z_2^t Z_2| Q(\gamma_1) \\
 &= 2N\sigma^{-2k} |Z_2^t Z_2| Q(\rho)
 \end{aligned} \tag{4.3.8}$$

Ahora bien, de acuerdo con (4.1.11) se tiene que la densidad marginal inicial de referencia para $\rho = \gamma_1$ está determinada por la relación

$$\pi(\rho) \propto \{ |I(\gamma^0)| / |H| \}^{1/2}$$

así que de (4.3.7) y (4.3.8) se tiene

$$\begin{aligned}
 \pi(\rho) &\propto \{ [2N\sigma^{-2(k+1)} |Z^t Z| \gamma_2^2] / [2N\sigma^{-2k} |Z_2^t Z_2| Q(\rho)] \}^{1/2} \\
 &= \{ [\sigma^{-2} |Z^t Z| \gamma_2^2] / [|Z_2^t Z_2| Q(\rho)] \}^{1/2} \\
 &\propto [Q(\rho)]^{-1/2}
 \end{aligned}$$

como función de ρ . Por otra parte, de (4.1.12) la densidad condicional inicial para $\gamma_2, \dots, \gamma_k$, o dado ρ está definida por la relación

$$\begin{aligned}
 \pi(\gamma_2, \dots, \gamma_k, \rho) &\propto \{ |H| \}^{1/2} \\
 &= \{ 2N\sigma^{-2k} |Z_2^t Z_2| Q(\rho) \}^{1/2}
 \end{aligned}$$

$$\propto \sigma^{-k}$$

como función de $\gamma_2, \dots, \gamma_k, \sigma$. En consecuencia, se tiene la siguiente

Proposición 1. La densidad inicial conjunta de referencia para el vector χ^* , cuando el parámetro de interés es $\rho = \gamma_1$, está determinada por la relación

$$\begin{aligned} \pi(\rho, \gamma_2, \gamma_3, \dots, \gamma_k, \sigma) &= \pi(\gamma_2, \gamma_3, \dots, \gamma_k, \sigma | \rho) \pi(\rho) \\ &\propto [Q(\rho)]^{-1/2} \sigma^{-k}. \end{aligned}$$

Un resultado inmediato es el siguiente

Corolario 1. La densidad inicial conjunta de referencia para χ^* , cuando el parámetro de interés es $\rho = \gamma_1$, pertenece a la familia propuesta en la sección 3.2.

Demostración. Recurriendo a la expresión (3.2.5) basta con tomar $r = k$ e identificar

$$p(\rho) = \pi(\rho)$$

$$\propto [Q(\rho)]^{-1/2}. \quad \square$$

Una consecuencia notable de este resultado es que la densidad inicial marginal de referencia para el parámetro de interés, $[Q(\rho)]^{-1/2}$ es propia (Gradshteyn y Rhizik 1965, página 81) aunque la conjunta resulta impropia.

Más aún, del Teorema 1 de la sección 3.3 se tiene que la correspondiente densidad final marginal de referencia para ρ está

dada por

$$\pi(\rho | \underline{y}) \propto [Q(\rho)]^{-1} [H(\rho)]^{-N/2} \quad (4.3.9)$$

y es propia puesto que $p(\rho) = [Q(\rho)]^{-1/2}$ es propia como ya se ha indicado. Adicionalmente del Teorema 1 de la sección 3.5 se puede deducir que $\pi(\rho | \underline{y})$ tiene la expresión alternativa

$$\pi(\rho | \underline{y}) \propto [Q(\rho)]^{-1} [\eta + h(\rho)]^{-1/2} \quad (4.3.10)$$

en donde $\eta = N-k$ y

$$h(\rho) = (\hat{\beta}_1 - \rho \hat{\beta}_2)^2 / [S^2 \{v_{22} \rho^2 - 2v_{12} \rho + v_{11}\}].$$

Finalmente, es posible caracterizar en mayor detalle la densidad final conjunta de referencia $\pi(\rho | \underline{y})$. De las expresiones (4.3.9) y (4.3.10) es claro que $H(\rho)$ es un cociente de dos polinomios de grado dos en ρ . Más aún, de (3.3.33) se tiene que

$$H(\rho) = J(\rho) / Q(\rho) \quad (4.3.11)$$

con

$$\begin{aligned} J(\rho) &= Q(\rho) S_Y^2 - [S_{ZY2} + \rho S_{ZY1}]^2 \\ &= Q(\rho) S_Y^2 - S_{ZY2}^2 - 2 S_{ZY1} S_{ZY2} \rho - S_{ZY1}^2 \rho^2 \\ &= [c_2 \rho^2 + c_1 \rho + c_0] S_Y^2 \\ &\quad - S_{ZY2}^2 - 2 S_{ZY1} S_{ZY2} \rho - S_{ZY1}^2 \rho^2 \\ &= [c_2 S_Y^2 - S_{ZY1}^2] \rho^2 + [c_1 S_Y^2 - 2 S_{ZY1} S_{ZY2}] \rho \end{aligned}$$

$$\begin{aligned}
 & + [c_0 S_Y^2 - S_{ZY_2}^2] \\
 & = \delta_2 \rho^2 + \delta_1 \rho + \delta_0
 \end{aligned}
 \tag{4.3.12}$$

en donde por supuesto

$$\begin{aligned}
 \delta_2 & = [c_2 S_Y^2 - S_{ZY_1}^2] \\
 \delta_1 & = [c_1 S_Y^2 - 2 S_{ZY_1} S_{ZY_2}] \\
 \delta_0 & = [c_0 S_Y^2 - S_{ZY_2}^2].
 \end{aligned}$$

De esta manera y dado que $\pi(\rho | \underline{Y})$ es de clase C^∞ en \mathbb{R} y es estrictamente positiva para todo real ρ , se sigue que

$$\begin{aligned}
 \log\{\pi(\rho | \underline{Y})\} & = \log\{[Q(\rho)]^{-1} [H(\rho)]^{-N/2}\} \\
 & = -\log\{Q(\rho)\} - (N/2)\log\{H(\rho)\} \\
 & = -\log\{Q(\rho)\} - (N/2)[\log\{J(\rho)\} - \log\{Q(\rho)\}] \\
 & = (N/2 - 1)\log\{Q(\rho)\} - (N/2)\log\{J(\rho)\}
 \end{aligned}$$

de donde

$$\frac{\partial \log\{\pi(\rho | \underline{Y})\}}{\partial \rho} = (N/2 - 1) \frac{Q'(\rho)}{Q(\rho)} - (N/2) \frac{J'(\rho)}{J(\rho)}$$

... (4.3.14)

y en consecuencia, los valores extremos de $\pi(\rho | \underline{Y})$ pueden obtenerse de la ecuación

$$(N/2 - 1) \frac{Q'(\rho)}{Q(\rho)} - (N/2) \frac{J'(\rho)}{J(\rho)} = 0
 \tag{4.3.14}$$

o equivalentemente

$$(1 - N/2) J(\rho)Q'(\rho) + (N/2) Q(\rho)J'(\rho) = 0 \quad (4.3.15)$$

Puesto que tanto Q con J son polinomios de grado dos, las correspondientes derivadas Q' y J' son polinomios de grado uno y entonces los productos $J'Q$ y $Q'J$ resultan polinomios de grado tres. De esta manera, un valor de ρ anula la derivada (4.3.13) si y sólo si satisface la ecuación de grado a lo más tres (4.3.15). Ahora bien, de (4.3.12) se tiene que

$$J(\rho) = \delta_2 \rho^2 + \delta_1 \rho + \delta_0$$

de forma que

$$J'(\rho) = 2\delta_2 \rho + \delta_1 \quad (4.3.16)$$

Análogamente,

$$Q(\rho) = c_2 \rho^2 + c_1 \rho + c_0$$

y por lo tanto

$$Q'(\rho) = 2c_2 \rho + c_1 \quad (4.3.17)$$

Entonces,

$$\begin{aligned} JQ' &= (\delta_2 \rho^2 + \delta_1 \rho + \delta_0)(2c_2 \rho + c_1) \\ &= (2\delta_2 c_2) \rho^3 + (2\delta_1 c_2) \rho^2 + (2\delta_0 c_2) \rho \\ &\quad + (\delta_2 c_1) \rho^2 + (\delta_1 c_1) \rho + (\delta_0 c_1) \end{aligned}$$

$$\begin{aligned}
 &= (2c_2 c_2) \rho^3 + (2c_1 c_2 + c_2 c_1) \rho^2 + (2c_0 c_2 + c_1 c_1) \rho \\
 &+ (c_0 c_1) \quad (4.3.18)
 \end{aligned}$$

$$\begin{aligned}
 QJ' &= (c_2 \rho^2 + c_1 \rho + c_0)(2c_2 \rho + c_1) \\
 &= (2c_2 c_2) \rho^3 + (2c_1 c_2 + c_2 c_1) \rho^2 + (2c_0 c_2 + c_1 c_1) \rho \\
 &+ (c_0 c_1) \quad (4.3.19)
 \end{aligned}$$

de donde

$$(1 - N/2)J(\rho)Q'(\rho) + (N/2)Q(\rho)J'(\rho) = \omega_3 \rho^3 + \omega_2 \rho^2 + \omega_1 \rho + \omega_0$$

con

$$\begin{aligned}
 \omega_3 &= (1 - N/2)(2c_2 c_2) + (N/2)(2c_2 c_2) \\
 &= 2c_2 c_2 \\
 &= 2[c_2 S_Y^2 - S_{xy}^2]c_2
 \end{aligned}$$

De la observación que sigue a la definición del polinomio $Q(\rho)$ en la sección 3.3 - expresión (3.3.15) - se tiene que $c_2 > 0$, en consecuencia $\omega_3 = 0$ si y sólo si $c_2^* = 0$. Sin embargo, de (3.3.39) se tiene que $\text{pl } c^* = 0 \mid = 0$ con

$$c^* = S_Y^2 - S_{xy}^2 / c_2.$$

Por lo tanto, $c_2^* \neq 0$ con probabilidad uno y como consecuencia, la ecuación

$$\frac{\partial \pi(\rho | Y)}{\partial \rho} = 0 \quad (4.3.20)$$

es equivalente a una ecuación de grado exactamente tres. Puesto que un polinomio de grado tres sólo puede tener o bien una raíz real o bien sus tres raíces reales y dado que $\pi(\rho | Y)$ es propia y ρ toma valores en toda la recta real, se sigue la siguiente

Proposición 2. La densidad final marginal de referencia $\pi(\rho | Y)$ tiene a lo más dos modas. Es unimodal si la ecuación (4.3.20) tiene una única raíz real y es bimodal si (4.3.20) tiene tres raíces reales.

En los siguientes capítulos, los resultados obtenidos en este trabajo se ilustrarán con conjuntos de datos generados utilizando algunos casos particulares y bien conocidos del modelo general descrito en (1.1.1) y (1.1.2).

CAPITULO 5. EL MODELO DE COCIENTE DE PENDIENTES

5.1 El modelo.

En este capítulo se considerará la siguiente estructura, que es muy común en el análisis de ensayos biológicos. Suponga que se realiza un experimento en el que p dosis $\{ X_{11}, \dots, X_{1p} \}$ de un primer estímulo y q dosis $\{ X_{21}, \dots, X_{2q} \}$ de un segundo estímulo se ensayan n veces de manera que se obtiene un conjunto $\{ Y_{ijk} : j=1, \dots, p ; k=1, \dots, n ; Y_{ijk} : j=1, \dots, q ; k=1, \dots, n \}$ de observaciones condicionalmente independientes, con distribución Normal y varianza común σ^2 tales que

$$\begin{aligned} E(Y_{ijk}) &= \alpha + \beta X_{1j} & ; j=1, \dots, p; k=1, \dots, n \\ E(Y_{ijk}) &= \alpha + \rho\beta X_{2j} & ; j=1, \dots, q; k=1, \dots, n \end{aligned} \quad (5.1.1)$$

Así, Y_{ijk} representa la k -ésima medición de la variable de respuesta Y asociada a la aplicación de X_{1j} , la j -ésima dosis del i -ésimo estímulo.

Si las dosis X_1 del primer estímulo y X_2 del segundo estímulo son equivalentes, esto es si las correspondientes respuestas esperadas son iguales [$E(Y_1) = E(Y_2)$], entonces su

potencia relativa se define simplemente como el cociente X_1 / X_2 . Utilizando la estructura a (5.1.1) es inmediato que

$$\begin{aligned} X_1 / X_2 &= \rho \beta / \beta \\ &= \rho. \end{aligned}$$

Este valor existe si β es distinto de cero y es independiente de la particular pareja de dosis equivalentes como consecuencia de la ordenada común en (5.1.1). Esta característica estructural del modelo aparece al adoptar la llamada *condición de similaridad* (Finney 1952/78, sección 3.3) para los estímulos y conduce, como se ha descrito, a la situación en que el problema de establecer la potencia relativa se convierte en el problema de estimar el cociente de pendientes en un conjunto de dos modelos de regresión lineal simple con ordenada común. Esta estructura se conoce en la literatura, por motivo obvios, con el nombre de *ensayo de cociente de pendientes* (Bliss 1970, Armitage 1977) y produce, para un conjunto de datos D, la función de verosimilitud

$$\begin{aligned} L(p, \alpha, \beta, \sigma^2 | D) &= \alpha \sigma^{-n(p+q)} \exp \left\{ - \left[\sum_{j=1}^p \sum_{k=1}^n (Y_{1jk} - \alpha - \beta X_{1j})^2 \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^q \sum_{k=1}^n (Y_{2jk} - \alpha - \rho \beta X_{2j})^2 \right] / (2\sigma^2) \right\} \\ &\dots (5.1.3) \end{aligned}$$

El modelo que describe el ensayo de cociente de pendientes es un caso particular del modelo (1.1.1). Basta establecer las siguientes relaciones

$$\underline{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ \vdots \\ Y_{1pn} \\ Y_{211} \\ Y_{212} \\ \vdots \\ Y_{2qn} \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & 0 \\ 1 & X_{11} & 0 \\ \vdots & \vdots & \vdots \\ 1 & X_{1p} & 0 \\ 1 & 0 & X_{21} \\ 1 & 0 & X_{21} \\ \vdots & \vdots & \vdots \\ 1 & 0 & X_{2q} \end{bmatrix}$$

... (S.1.4)

Esto es, el vector $\underline{Y} \in \mathbb{R}^N$; $N = n(p+q)$ se obtiene de ordenar lexicográficamente las respuestas observadas y la matriz X ($N \times 3$) tiene la primera columna con todas sus entradas iguales a uno, la segunda columna con la últimas nq entradas iguales a cero y las primeras np agrupadas en p bloques de tamaño n de manera que todas las entradas del j -ésimo bloque son iguales a X_{1j} ; por su parte, la tercera columna de X tiene las np primeras entradas iguales a cero y las restantes nq agrupadas en q bloques de tamaño n de forma que todas las entradas del j -ésimo bloque son iguales a X_{2j} . Si se definen

$$\underline{\theta} = \begin{bmatrix} \alpha \\ \beta \\ \beta \end{bmatrix} \quad \underline{\lambda}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \underline{\lambda}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad (\text{S.1.4})$$

es claro que

$$\underline{Y} \sim NC \underline{X} \underline{\theta}, \sigma^2 I$$

y además

$$\rho = \frac{\lambda_1^t \theta}{\lambda_2^t \theta}$$

Ahora bien, para pasar al modelo reparametrizado (4.3.1) es necesario construir la matriz L tal que

$$L^t = \langle \lambda_1, \lambda_2, \lambda_3 \rangle$$

sea de rango completo. En este caso, puesto que $\lambda_1 = \underline{e}_3$ y $\lambda_2 = \underline{e}_2$, con \underline{e}_i el i -ésimo vector canónico en \mathbb{R}^3 , es natural la selección $\lambda_3 = \underline{e}_1$. De esta forma,

$$L = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

y entonces

$$L^{-1} = L$$

de donde se sigue que si $Z = X L^{-1}$, entonces

$$Z = \begin{bmatrix} 1 & X_{11} & 0 \\ 1 & X_{11} & 0 \\ \vdots & \vdots & \vdots \\ 1 & X_{1p} & 0 \\ 1 & 0 & X_{21} \\ 1 & 0 & X_{21} \\ \vdots & \vdots & \vdots \\ 1 & 0 & X_{2q} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & X_{11} & 1 \\ 0 & X_{21} & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & X_{1p} & 1 \\ X_{21} & 0 & 1 \\ X_{21} & 0 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ X_{2q} & 0 & 1 \end{bmatrix}$$

Esto es, en términos de columnas, $Z = (Z_1, Z_2, Z_3)$ con

$$Z_1^t = (0, 0, \dots, 0, X_{21}, X_{21}, \dots, X_{2q})$$

$$Z_2^t = (X_{11}, X_{11}, \dots, X_{1p}, 0, 0, \dots, 0)$$

$$Z_3^t = (1, 1, \dots, 1, 1, 1, \dots, 1)$$

En forma análoga se tiene que si $\beta = L \theta$, entonces

$$\begin{aligned} \theta &= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \rho\beta \end{bmatrix} \\ &= \begin{bmatrix} \rho\beta \\ \beta \\ \alpha \end{bmatrix} \end{aligned}$$

y se obtiene el modelo reparametrizado

$$\underline{Y} \sim NC \underline{Z}\beta, \sigma^2 I$$

y la reexpresión de el parámetro de interés $\rho = \beta_1 / \beta_2$.

5.2 La distribución final de referencia para ρ .

Ahora, utilizando la expresión (4.3.9) para la densidad final marginal de referencia para ρ y la expresión (4.3.11) para $H(\rho)$ se tiene que

$$\begin{aligned} \pi(\rho | \underline{Y}) &\propto [Q(\rho)]^{-1} [H(\rho)]^{-N/2} \\ &= [Q(\rho)]^{-1} [J(\rho) / Q(\rho)]^{-N/2} \\ &= [Q(\rho)]^{-1} [S_Y^2 - \{S_{ZY2} + \rho S_{ZY1}\}^2 / Q(\rho)]^{-N/2} \end{aligned}$$

... (5.2.1)

en donde

$$S_Y^2 = \underline{Y}^t (I - P) \underline{Y}$$

$$S_{ZY1} = \underline{Z}_1^t (I - P) \underline{Y}$$

$$S_{ZY2} = \underline{Z}_2^t (I - P) \underline{Y}$$

y $Q(\rho) = c_2 \rho^2 + c_1 \rho + c_0$, con

$$c_2 = \underline{Z}_1^t (I - P) \underline{Z}_1$$

$$c_1 = 2 \mathbb{Z}_1^t (I - P) \mathbb{Z}_2^t$$

$$c_0 = \mathbb{Z}_2^t (I - P) \mathbb{Z}_2^t$$

y la matriz $(I - P)$ definida, como en (3.3.13), por

$$\begin{aligned} I - P &= I - \mathbb{Z}_2 (\mathbb{Z}_2^t \mathbb{Z}_2)^{-1} \mathbb{Z}_2^t \\ &= I - \mathbb{Z}_0 (\mathbb{Z}_0^t \mathbb{Z}_0)^{-1} \mathbb{Z}_0^t \\ &= I - \mathbb{Z}_0 (N)^{-1} \mathbb{Z}_0^t \\ &= I - (\mathbb{Z}_0 \mathbb{Z}_0^t) / N \\ &= I - M / N \end{aligned}$$

con M la matriz $(N \times N)$ cuyas entradas son todas iguales a 1. Por lo tanto, $I - M / N$ es la matriz que transforma cada vector en el correspondiente vector de residuos respecto a la media de las entradas. De esta forma,

$$S_Y^2 = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y})^2$$

$$S_{ZV_1} = \sum_j \sum_k (Y_{zjk} - \bar{Y}) X_{zj}$$

$$S_{ZV_2} = \sum_j \sum_k (Y_{ijk} - \bar{Y}) X_{ij}$$

con $\bar{Y} = (\sum_i \sum_j \sum_k Y_{ijk}) / N$. Además,

$$c_2 = n \sum_j X_{zj} - n^2 (\sum_j X_{zj})^2 / N$$

$$c_1 = -2 (n^2 / N) (\sum_j X_{ij}) (\sum_j X_{zj})$$

$$c_0 = n \sum_j x_{1j} - n^2 (\sum_j x_{1j})^2 / N$$

con lo cual, a partir de los datos originales es posible calcular el núcleo de la densidad $\pi(\rho | \underline{Y})$.

5.3 Ejemplos numéricos.

Con el propósito de ilustrar los resultados que se pueden obtener con la densidad final de referencia, $\pi(\rho | \underline{Y})$ ha sido calculada para tres distintos conjuntos de datos y comparada con el estimador de máxima verosimilitud de ρ y la región de confianza que se obtiene via el Teorema de Fieller. El primer conjunto consta de datos reales, procede de un ensayo reportado inicialmente por Wood (1964) y analizado por Finney (1978, página 161) y se exhibe en la Tabla 1.

TABLA 1.
Datos de Wood ($n = 4$, $p = 3$, $q = 2$)

Estímulo Dosis	1			2	
	0.0	0.5	1.0	0.5	1.0
38	097	167	80	121	
45	100	164	88	124	
40	105	159	90	122	
44	098	156	82	122	

Para este conjunto de datos, el estimador de máxima verosimilitud de ρ obtenido por Finney es $r = 0.6847$ y el intervalo del 95% de confianza para ρ resulta (0.6464, 0.7235). Es claro de la Figura 1 que la final $\pi(\rho | \underline{Y})$ es consistente con estos resultados. De hecho, la única moda de esta densidad se localiza en $\rho = 0.6846$ y la probabilidad que esta densidad asigna

al intervalo (0.6464, 0.7235) es 0.9608.

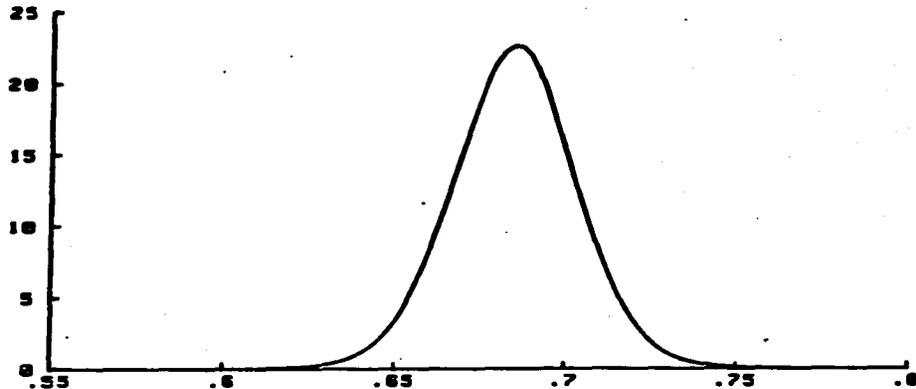


FIGURA 1. Final de referencia $\pi(p | \underline{y})$ para los datos de Wood.

Los siguientes dos ejemplos han sido obtenidos mediante simulación utilizando el modelo (5.1.1). El segundo conjunto, que se muestra en la Tabla 2, es un ejemplo de lo que podría llamarse un experimento "razonable" en el sentido de que ambos estímulos tienen un efecto apreciable en la respuesta.

TABLA 2.

Datos simulados ($n = 3$, $p = q = 4$, $\alpha = 0$, $\beta = 10$, $p\beta = 20$, $\sigma = 1$)

Estímulo Dosis	1				2			
	0.5	1.5	2.5	3.5	1.0	2.0	3.0	4.0
	5.784	16.479	25.948	33.770	21.301	39.818	60.333	81.391
	3.008	15.023	24.347	37.054	21.508	40.989	57.723	80.396
	5.508	14.594	25.228	33.831	21.047	38.550	59.430	81.443

En la Figura 2 se exhibe la parte principal de $\pi(\rho | \underline{y})$ en donde se puede observar una moda en $\rho = 2.009$ - muy cerca del verdadero valor $\rho = 2$ - . La masa de probabilidad está claramente acumulada alrededor de este punto puesto que se puede verificar que $\pi(1.91 < \rho < 2.09 | \underline{y}) = 0.954$ y si se ha indicado que esta es la parte principal de la densidad final es porque existe otra moda, localizada en $\rho = -14.721$. Sin embargo, debido a que $\pi(\rho < 0 | \underline{y}) < 10^{-10}$ esta segunda moda puede ser ignorada sin consecuencias prácticas.

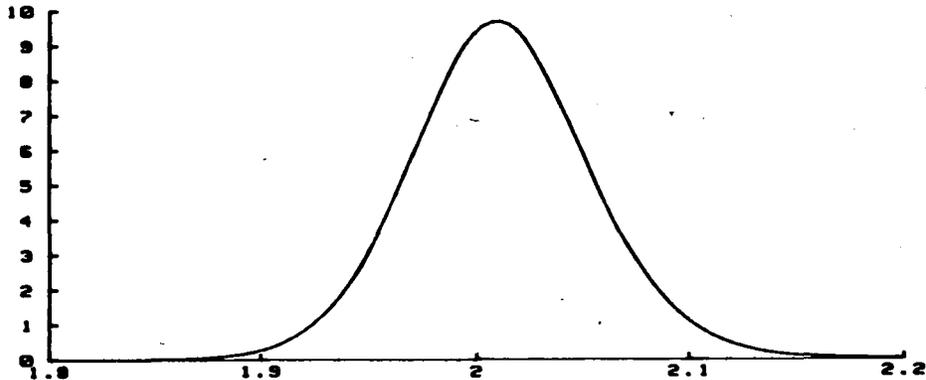


FIGURA 2. Final de referencia $\pi(\rho | \underline{y})$ datos de la Tabla 2.

En este caso nuevamente los resultados no difieren mayormente de los que se obtienen con el enfoque frecuentista. De hecho, el valor estimado por máxima verosimilitud es $\hat{\rho} = 2.011$ y el intervalo del 95% de confianza está dado por $(1.926, 2.104)$.

Finalmente, la Tabla 3 exhibe otro conjunto de datos

simulados. Estos datos fueron generados utilizando valores de los parámetros que conducen a una situación más extrema en donde la información provista por el experimento podría sugerir, si se analiza con los procedimientos frecuentistas usuales, que el primer estímulo no tiene un efecto apreciable en la respuesta.

TABLA 3.

Datos simulados ($n = 3, p = q = 4, a = 2, \beta = 0.3, p\beta = 1, \sigma = 1$)

Estímulo Dosis	0.5	1.5 ¹	2.5	3.5	1.0	2.0 ²	3.0	4.0
	2.612	2.600	2.509	2.398	2.700	2.714	4.519	4.530
	1.946	2.672	0.986	2.820	2.010	4.881	5.517	7.244
	2.803	2.001	2.533	2.209	2.024	6.023	6.090	5.297

La final de referencia $\pi(p | \underline{Y})$ describe, como se muestra en la

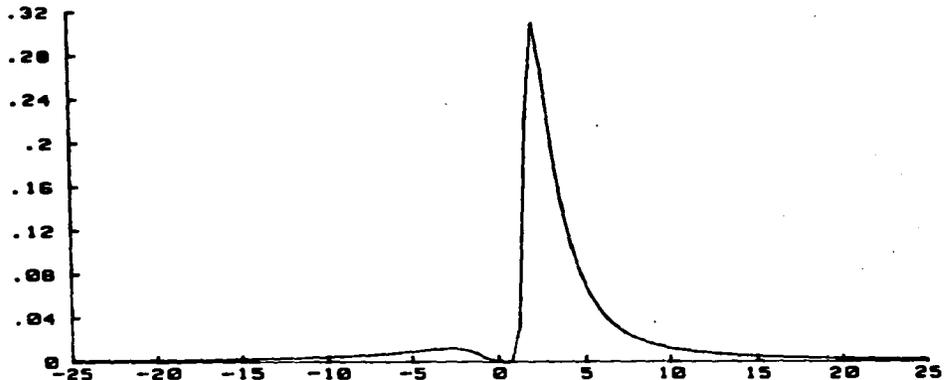


FIGURA 3. Final de referencia $\pi(p | \underline{Y})$ datos de la Tabla 3.

Figura 3. lo que se puede afirmar sobre el parámetro de interés. Por ejemplo, se puede verificar que $\pi(|\rho| < 1 | \underline{Y})$ es menor que 0.005 de manera que $|\rho| < 1$ es a posteriori, un evento muy poco probable.

Por otra parte, $\pi(\rho > 0 | \underline{Y}) = 0.875$ por lo que existe un apreciable grado de certeza respecto al signo del cociente. Más aun, $\pi(1 < \rho < 4 | \underline{Y}) = 0.550$ y $\pi(1 < \rho < 9 | \underline{Y}) = 0.770$ de donde se sigue que la densidad final de referencia es claramente informativa y compatible con el verdadero valor del parámetro $\rho = 10/3$. Si se aplican los procedimientos frecuentistas a este conjunto de datos se tiene que el valor del estimador por máxima verosimilitud está dado por $r = 3.622$ que podría considerarse una estimación razonable. Sin embargo, la región de confianza de nivel 95% que se obtiene utilizando el Teorema de Fieller esta dada por $(-\infty, -3.068) \cup (1.530, \infty)$. Los inconvenientes de una región de este tipo ya han sido comentados en la sección 2.4 y una observación relevante se puede encontrar en Finney (1952/78, página 82).

Un examen más detallado de la densidad bimodal y estrictamente positiva $\pi(\rho | \underline{Y})$ revela que si $p \in (0.783, 0.987)$ la correspondiente región de máxima densidad de tamaño p (rmd-p) consta de la unión de dos intervalos ajenos y finitos. Si p es menor que 0.783 o mayor que 0.987, la rmd-p respectiva está dada por un único intervalo finito ($p < 1$).

CAPITULO 6. EL MODELO DE LINEAS PARALELAS

6.1 El modelo.

En forma similar al tratamiento que se dió en el capítulo 5 al modelo de cociente de pendientes, en este capítulo se particularizará en otra estructura que muy comunmente se utiliza para describir los resultados de cierto tipo de bicensayos. En esta ocasion suponga que se realiza un experimento por medio del cual una colección de p dosis, estrictamente positivas, $\{ X_{11}, \dots, X_{1p} \}$ de un primer estímulo y q dosis, tambien estrictamente positivas, $\{ X_{21}, \dots, X_{2q} \}$ de un segundo estímulo se ensayan n veces de manera que se obtiene un conjunto $\{ Y_{ijk} ; j=1, \dots, p ; k=1, \dots, n ; Y_{ijk} ; j=1, \dots, q ; k=1, \dots, n \}$ de observaciones condicionalmente independientes, con distribución Normal y varianza comun σ^2 tales que

$$E(Y_{ijk}) = \alpha_1 + \beta \log(X_{1j}) \quad ; j=1, \dots, p; k=1, \dots, n$$

(6.1.1)

$$E(Y_{ijk}) = \alpha_2 + \beta \log(X_{2j}) \quad ; j=1, \dots, q; k=1, \dots, n$$

En este modelo Y_{ijk} representa, como en (6.1.1), la k -ésima medición de la variable de respuesta Y asociada a la

aplicación de X_{ij} , la j -ésima dosis del i -ésimo estímulo. Si de nuevo las dosis X_1 del primer estímulo y X_2 del segundo estímulo son equivalentes en el sentido de que las respuestas esperadas correspondientes son iguales [$E(Y_1) = E(Y_2)$], entonces el cociente X_1 / X_2 que define la potencia relativa puede obtenerse como sigue

$$E(Y_1) = \alpha_1 + \beta \log(X_1)$$

$$E(Y_2) = \alpha_2 + \beta \log(X_2)$$

de donde, si $\beta \neq 0$, se tiene

$$\log(X_1) = (E(Y_1) - \alpha_1) / \beta$$

$$\log(X_2) = (E(Y_2) - \alpha_2) / \beta$$

Puesto que $E(Y_1) = E(Y_2)$, se sigue que si se define la magnitud $\psi = \log(X_1 / X_2)$, entonces

$$\psi = \log(X_1) - \log(X_2)$$

$$= (\alpha_2 - \alpha_1) / \beta$$

(5.1.2)

y finalmente $\rho = \exp(\psi)$. De esta manera, la potencia relativa es una función uno-a-uno de ψ que a su vez es un parámetro de la forma que se trata en esta tesis. Más aún, ψ existe si β es distinto de cero y resulta independiente de la particular pareja de dosis equivalentes como consecuencia de la pendiente común en (5.1.1). Este modelo, conocido como de líneas paralelas, es de uso muy frecuente y se encuentra documentado por ejemplo en el libro de Finney (1952/78). Dada la relación entre los parámetros ψ y ρ , es posible aprovechar las propiedades de las distribuciones de referencia (Bernardo, 1979) para obtener la final de referencia

para ρ a partir de la correspondiente final de referencia para y aplicando el cambio de variable respectivo. De esta forma, en lo que resta se reexpresara el modelo (6.1.1) como sigue

$$E(Y_{1jk}) = \alpha + \beta U_{1j} \quad ; \quad j=1, \dots, p; \quad k=1, \dots, n \quad (6.1.3)$$

$$E(Y_{2jk}) = \alpha + \beta (U_{2j} + \psi) \quad ; \quad j=1, \dots, q; \quad k=1, \dots, n$$

en donde $U_{ij} = \log(X_{ij})$ de manera que para un conjunto de datos D se obtiene la siguiente función de verosimilitud

$$L(\psi, \alpha, \beta, \sigma | D) = \alpha \sigma^{-n(p+q)} \exp\left\{-\left[\sum_{j=1}^p \sum_{k=1}^n (Y_{1jk} - \alpha - \beta U_{1j})^2 + \sum_{j=1}^q \sum_{k=1}^n (Y_{2jk} - (\alpha + \beta\psi) - \beta U_{2j})^2\right] / (2\sigma^2)\right\} \quad \dots(6.1.4)$$

Este modelo es también un caso particular del modelo (1.1.1). Como en el capítulo 5, basta con identificar el vector \underline{Y} , la matriz X y el vector $\underline{\theta}$.

$$\underline{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ \vdots \\ Y_{1pn} \\ Y_{211} \\ Y_{212} \\ \vdots \\ Y_{2qn} \end{bmatrix} \quad X = \begin{bmatrix} 1 & 0 & U_{11} \\ 1 & 0 & U_{11} \\ \vdots & \vdots & \vdots \\ 1 & 0 & U_{1p} \\ 1 & 1 & U_{21} \\ 1 & 1 & U_{21} \\ \vdots & \vdots & \vdots \\ 1 & 1 & U_{2q} \end{bmatrix} \quad (6.1.5)$$

Nuevamente $\underline{Y} \subset \mathbb{R}^N$; $N = n(p+q)$) es el vector que se obtiene al disponer en orden lexicográfico las observaciones producidas por el experimento mientras que la matriz X ($N \times 3$) tiene una primer columna tal que todas sus entradas son iguales a uno, la segunda columna tiene sus primeras np entradas iguales a cero con las últimas nq entradas iguales a uno; finalmente, la tercera columna tiene sus entradas agrupadas en $p+q$ bloques de tamaño n de manera que todas las entradas del j -ésimo bloque son iguales a U_{1j} si $j \in \{1, \dots, p\}$ y son iguales a U_{2j} si $j \in \{p+1, \dots, p+q\}$. Si además se definen

$$\underline{\theta} = \begin{bmatrix} \alpha \\ \beta\psi \\ \beta \end{bmatrix} \quad \lambda_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \lambda_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (8.1.0)$$

se tiene entonces que

$$\underline{Y} \sim NC X\underline{\theta}, \sigma^2 I)$$

y además

$$w = \lambda_1^t \underline{\theta} / \lambda_2^t \underline{\theta}$$

Exactamente igual que en el caso del cociente de pendientes, para obtener el modelo reparametrizado (4.3.1) es necesario construir la matriz de rango completo L tal que

$$L^t = (\lambda_1, \lambda_2, \lambda_3)$$

Puesto que $\lambda_1 = \underline{e}_2$ y $\lambda_2 = \underline{e}_3$, con \underline{e}_i el i -ésimo vector canónico en \mathbb{R}^3 , $\lambda_3 = \underline{e}_1$ resulta la elección natural de forma que

$$L = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

y entonces

$$L^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

de donde si $Z = X L^{-1}$, se obtiene

$$Z = \begin{bmatrix} 1 & 0 & U_{11} \\ 1 & 0 & U_{11} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 1 & 0 & U_{1p} \\ 1 & 1 & U_{21} \\ 1 & 1 & U_{21} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 1 & 1 & U_{2q} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & U_{11} & 1 \\ 0 & U_{11} & 1 \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 0 & U_{1p} & 1 \\ 1 & U_{21} & 1 \\ 1 & U_{21} & 1 \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 1 & U_{2q} & 1 \end{bmatrix}$$

Esto es, la matriz Z se puede expresar, en términos de

vectores columnas como $Z = (z_1, z_2, z_3)$ con

$$\Xi_1^t = (0, 0, \dots, 0, 1, 1, \dots, 1)$$

$$\Xi_2^t = (U_{11}, U_{11}, \dots, U_{1p}, U_{21}, U_{21}, \dots, U_{2q})$$

$$\Xi_3^t = (1, 1, \dots, 1, 1, 1, \dots, 1)$$

En forma análoga se tiene que si $\beta = L \theta$, entonces

$$\begin{aligned} \beta &= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \beta \end{bmatrix} \\ &= \begin{bmatrix} \beta \\ \beta \\ \alpha \end{bmatrix} \end{aligned}$$

de donde se obtiene el modelo reparametrizado

$$Y \sim NC(Z\beta, \sigma^2 I)$$

y la reexpresión de el parámetro w

$$w = \beta_1 / \beta_2$$

5.2 La distribución final de referencia para ψ .

Nuevamente recurriendo a la expresión (4.3.9) para la densidad final marginal de referencia para el cociente y la correspondiente expresión (4.3.11) para $H(\psi)$ se tiene que

$$\begin{aligned} \pi(\psi | \underline{Y}) &\propto [Q(\psi)]^{-1} [H(\psi)]^{-N/2} \\ &= [Q(\psi)]^{-1} [J(\psi) / Q(\psi)]^{-N/2} \\ &= [Q(\psi)]^{-1} [S_Y^2 - \{S_{ZY2} + \psi S_{ZY1}\}^2 / Q(\psi)]^{-N/2} \\ &\dots(6.2.1) \end{aligned}$$

en donde

$$S_Y^2 = \underline{Y}^t (I - P) \underline{Y}$$

$$S_{ZY1} = \underline{Z}_1^t (I - P) \underline{Y}$$

$$S_{ZY2} = \underline{Z}_2^t (I - P) \underline{Y}$$

y además $Q(\psi) = c_2 \psi^2 + c_1 \psi + c_0$, con

$$c_2 = \underline{Z}_1^t (I - P) \underline{Z}_1$$

$$c_1 = 2 \underline{Z}_1^t (I - P) \underline{Z}_2$$

$$c_0 = \underline{Z}_2^t (I - P) \underline{Z}_2$$

La matriz $(I - P)$ por su parte, se define como en (3.3.13)

$$I - P = I - Z_2 (Z_2^t Z_2)^{-1} Z_2^t$$

$$\begin{aligned}
 &= I - \Xi_0 (\Xi_0^t \Xi_0)^{-1} \Xi_0^t \\
 &= I - \Xi_0 (N)^{-1} \Xi_0^t \\
 &= I - (\Xi_0 \Xi_0^t) / N
 \end{aligned}
 \tag{6.2.2}$$

Puesto que $\Xi_0^t = (1, 1, \dots, 1, 1, 1, \dots, 1)$, se tiene que

$$I - P = I - M / N \tag{6.2.3}$$

con M la matriz (N:N) cuyas entradas son todas iguales a 1. Por lo tanto se sigue, como en la sección 5.2, que

$$\begin{aligned}
 S_Y^2 &= \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y})^2 \\
 S_{ZV_1} &= \sum_j \sum_k (Y_{zjk} - \bar{Y}) \\
 S_{ZV_2} &= \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}) U_{ij}
 \end{aligned}$$

con $\bar{Y} = (\sum_i \sum_j \sum_k Y_{ijk}) / N$. Además,

$$\begin{aligned}
 c_2 &= nq - N (nq / N)^2 \\
 &= nq - n^2 q^2 / N \\
 &= nq [1 - nq / n(p+q)] \\
 &= nq [1 - q / (p+q)] \\
 &= (npq) / (p+q) \\
 c_1 &= 2n \sum_j U_{zj} - 2qn (nU_{..}) / N
 \end{aligned}$$

$$\begin{aligned}
 &= 2n \{ \sum_j U_{2j} - q\bar{U}_{..} \} \\
 &= 2nq \{ \bar{U}_2 - \bar{U}_{..} \} \\
 &= 2nq (p / (p+q)) \{ \bar{U}_2 - \bar{U}_1 \} \\
 &= (2npq / (p+q)) \{ \bar{U}_2 - \bar{U}_1 \}
 \end{aligned}$$

con

$$\begin{aligned}
 \bar{U}_1 &= (\sum_j U_{1j}) / p \\
 \bar{U}_2 &= (\sum_j U_{2j}) / q \\
 \bar{U}_{..} &= (p\bar{U}_1 + q\bar{U}_2) / (p + q)
 \end{aligned}$$

Finalmente.

$$\begin{aligned}
 c_o &= n \{ \sum_j (U_{1j} - \bar{U}_{..})^2 + \sum_j (U_{2j} - \bar{U}_{..})^2 \} \\
 &= n \sum_j \sum_j (U_{ij} - \bar{U}_{..})^2.
 \end{aligned}$$

Así, a partir de los datos originalmente producidos por el experimento, es posible calcular el núcleo de la densidad final $\pi(\psi | \underline{Y})$.

6.3 La distribución final de referencia para ρ .

La densidad $\pi(\psi | \underline{Y})$ tiene todas las propiedades que en general se discutieron en la sección 4.3. En particular, es propia y tiene a lo más dos modas. Si se requiere la densidad final de referencia para ρ , esta se puede obtener mediante el apropiado cambio de variable. Puesto que $\psi = \log(\rho)$, se tiene que

$$\pi(\rho | \underline{Y}) = \rho^{-1} \pi(\psi | \underline{Y})$$

$$\propto \rho^{-1} [Q(\rho)]^{-1} [H(\rho)]^{-N/2}$$

$$= [\rho Q(\log \rho)]^{-1} \{ S_Y^2 - [S_{\Sigma Y} + S_{\Sigma Y} \log(\rho)]^2 / Q(\log \rho) \}^{-N/2}$$

... (6.3.1)

Las características que posee $\pi(\rho | \underline{Y})$ se pueden derivar de las correspondientes de $\pi(\psi | \underline{Y})$. Es claro que $\pi(\rho | \underline{Y})$ está bien definida para todo $\rho \in \mathbb{R}^+$, es propia y en cuanto a sus modas se tiene que en virtud de que

$$\pi(\rho | \underline{Y}) = \rho^{-1} \pi(\psi(\rho) | \underline{Y})$$

entonces

$$\frac{\partial}{\partial \rho} \pi(\rho | \underline{Y}) = -\rho^{-2} \pi(\psi(\rho) | \underline{Y}) + \rho^{-1} \frac{\partial}{\partial \rho} \pi(\psi(\rho) | \underline{Y})$$

$$= -\rho^{-2} \pi(\psi(\rho) | \underline{Y}) + \rho^{-2} \frac{\partial}{\partial \psi} \pi(\psi | \underline{Y})$$

$$= \rho^{-2} \left[\frac{\partial}{\partial \psi} \pi(\psi | \underline{Y}) - \pi(\psi(\rho) | \underline{Y}) \right]$$

... (6.3.2)

Por lo tanto, los valores criticos de $\pi(\rho | \underline{y})$ son los puntos que satisfacen la relación

$$\rho^{-2} \left[\frac{\partial}{\partial \psi} \pi(\psi | \underline{y}) - \pi(\psi | \underline{y}) \right] = 0 \quad (6.3.3)$$

pero dado que $\rho = \exp(\psi) > 0$ se sigue que (6.3.3) es equivalente a

$$\frac{\partial}{\partial \psi} \pi(\psi | \underline{y}) - \pi(\psi | \underline{y}) = 0 \quad (6.3.4)$$

Más aún, debido a que la densidad $\pi(\psi | \underline{y})$ es estrictamente positiva para todo valor de $\psi \in \mathbb{R}$, (6.3.4) se puede transformar en

$$\left[\frac{\partial}{\partial \psi} \pi(\psi | \underline{y}) \right] / \pi(\psi | \underline{y}) - 1 = 0$$

o equivalentemente

$$\frac{\partial [\log \pi(\psi | \underline{y})]}{\partial \psi} = 1. \quad (6.3.5)$$

Ahora bien, recurriendo a la expresión (4.3.13) se tiene que

$$\frac{\partial [\log \pi(\psi | \underline{y})]}{\partial \psi} = (N/2 - 1) \frac{Q'(\psi)}{Q(\psi)} - (N/2) \frac{J'(\psi)}{J(\psi)} \quad \dots (6.3.6)$$

en donde tanto $Q(\psi)$ como $J(\psi)$ son polinomios de grado dos en ψ . En consecuencia, la relación

$$(N/2 - 1) \frac{Q'}{Q} - (N/2) \frac{J'}{J} = 1$$

que se puede reexpresar como

$$QJ - (N/2 - 1) JQ' + (N/2) QJ' = 0 \quad (6.3.7)$$

se satisface para todo valor de ψ que sea raíz del polinomio grado exactamente cuatro

$$G(\psi) = Q(\psi)J(\psi) - (N/2 - 1) J(\psi) Q'(\psi) + (N/2) Q(\psi)J'(\psi)$$

Puesto que ρ es una transformación uno-a-uno de ψ , se sigue que los valores críticos de $\pi(\rho | \underline{Y})$ se pueden obtener aplicando la transformación $\rho = \exp(\psi)$ a cada una de las raíces de $G(\psi)$. Por otra parte y dado que $G(\psi)$ es un polinomio de grado cuatro, se concluye que $\pi(\rho | \underline{Y})$ tiene cero, dos o bien cuatro puntos críticos. Ahora bien en virtud de que por construcción del modelo todas las dosis son positivas, el parámetro ρ sólo puede tomar valores en \mathbb{R}^+ y en consecuencia puede ocurrir que la densidad $\pi(\rho | \underline{Y})$ no tenga moda alguna, tenga exactamente una moda o que tenga dos modas. En cualquier caso es, a lo más, bimodal. Varios ejemplos que ilustran algunas de estas posibilidades se incluyen en la siguiente sección.

6.4 Ejemplos numéricos.

El primer ejemplo que se presenta en esta sección, como en la sección 5.3, consta de datos reales. Particularmente este conjunto de datos está referido a un ensayo de vitamina B12 que originalmente fue reportado por Emery et al. (1951) y que posteriormente ha sido analizado por Finney (1952/78, página 105). Los datos concretos se exhiben en la Tabla 1. En términos del parámetro de log-potencia ψ , se puede comprobar directamente que el estimador de máxima verosimilitud está dado por $\hat{\psi} = -1.5408$ mientras que el intervalo de confianza al 95% para ψ resulta $(-1.6835, -1.3915)$.

TABLA 1.
 Datos de Emery et al. ($n = 6$, $p = 3$, $q = 3$)

Estímulo	1			2		
Dosis	0.8	1.2	1.6	4	6	8
	0.96	1.06	1.17	0.91	1.09	1.15
	0.91	1.07	1.14	0.93	1.04	1.15
	0.92	0.99	1.14	0.98	0.97	1.14
	0.76	0.86	1.13	0.96	1.06	1.16
	1.03	1.06	1.13	0.89	1.04	1.10
	0.93	1.02	1.15	1.01	1.02	1.15

Por otra parte, si se calcula la densidad final marginal de referencia $\pi(\psi | \underline{Y})$ cuya gráfica se muestra en la Figura 1, se tiene que esta densidad tiene una única moda localizada en el

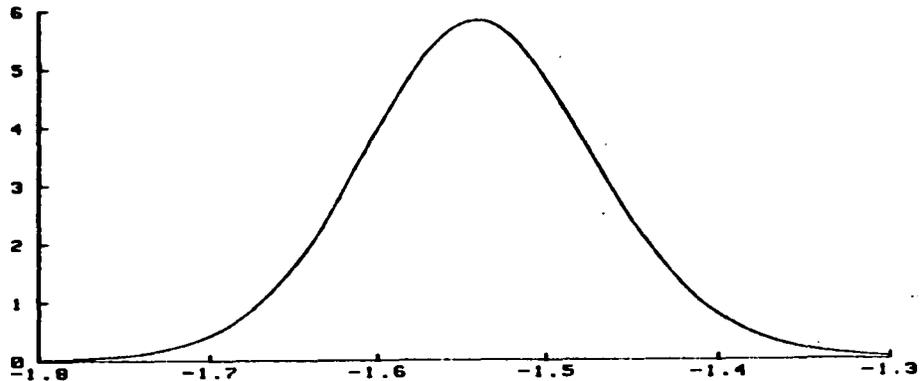


FIGURA 1. Final de referencia $\pi(\psi | \underline{Y})$ para los datos de Emery.

valor $\psi = -1.542$ y la probabilidad asignada al intervalo $(-1.6835, -1.3915)$ es 0.956. En conclusión, hay consistencia entre los resultados frecuentistas y bayesianos en este caso. Si se pretende obtener inferencias directamente sobre la potencia relativa ρ , entonces desde la perspectiva frecuentista y dado que $\rho = \exp(\psi)$, el correspondiente estimador de máxima verosimilitud puede calcularse a partir de f y resulta $r = 0.214$ con un intervalo de confianza al 95% para ρ de $(0.186, 0.249)$ tal como reporta Finney. Por su parte, la densidad final $\pi(\rho | Y)$ es unimodal con moda en $\rho = 0.213$ y por supuesto, también asigna al intervalo $(0.196, 0.244)$ una probabilidad de 0.956. La parte más importante de esta densidad se exhibe en la Figura 2 y claramente es consistente con los resultados frecuentistas.

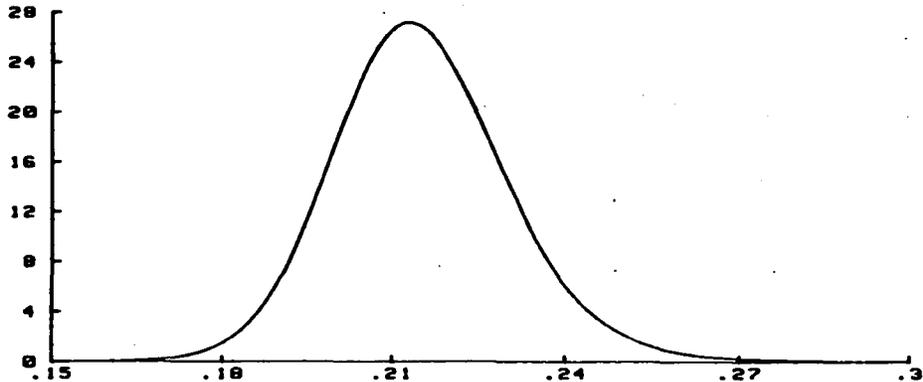


FIGURA 2. Final de referencia $\pi(\rho | Y)$ para los datos de Emery.

Si bien es cierto que la probabilidad $\pi(\rho < 0.05 | Y)$ es, para todos efectos prácticos, igual a cero resulta interesante

apuntar que $\pi(\rho | \underline{Y})$ tiene otro punto crítico - cf. discusión al final de la sección 6.3 - . Este punto crítico, un mínimo local para ser más precisos, se localiza en $\rho = 0.0132$ y el comportamiento cualitativo de $\pi(\rho | \underline{Y})$ en ese entorno se muestra en la Figura 3. Debe insistirse en que como consecuencia de que $\pi(\psi | \underline{Y})$ es propia, se sigue que $\pi(\rho | \underline{Y})$ es propia también aún cuando $\lim_{\rho \rightarrow 0} \pi(\rho | \underline{Y}) = \infty$.

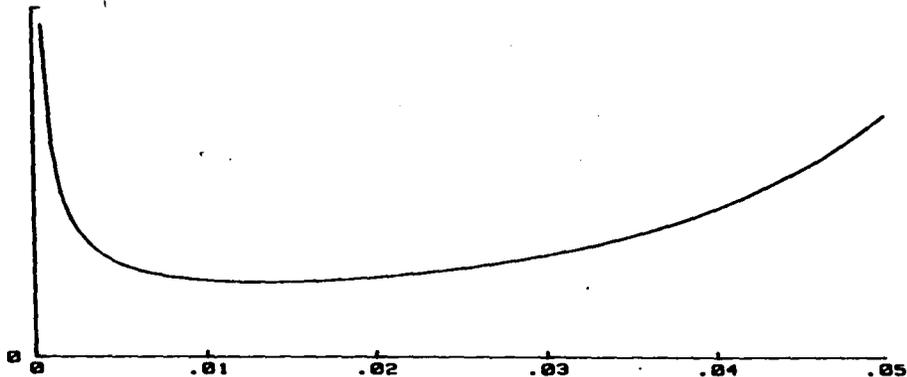


FIGURA 3. La densidad $\pi(\rho | \underline{Y})$ en un entorno de su mínimo local (datos de Emery).

El segundo ejemplo que se incluye en esta sección ha sido producido mediante simulación del modelo (6.1.3) y los datos correspondientes se muestran en la Tabla 2. En este caso y puesto que $\rho = 3$, se tiene que $\psi = \log(3) = 1.099$. Desde la perspectiva frecuentista y en relación a ψ , se tiene que la estimación de máxima verosimilitud resulta $f = 1.122$ con un intervalo de confianza al 95% para ψ de $(0.967, 1.286)$.

TABLA 2.

Datos simulados ($n = 5$, $p = q = 4$, $\alpha = 2$, $\beta = 5$, $\rho = 3$, $\sigma = 1$)

Estimulo	1				2			
Dosis	1	2	4	8	1	2	4	8
0.898	5.084	9.491	11.876	7.941	10.399	13.232	14.793	
2.582	5.351	10.250	12.033	10.131	9.720	13.292	18.808	
1.126	4.803	8.331	13.736	7.466	10.082	13.655	17.209	
2.571	5.270	9.042	10.308	6.977	10.845	15.083	18.910	
1.080	5.101	9.044	13.140	8.374	10.481	14.399	17.919	

Por otra parte, la densidad final de referencia $\pi(\psi | \underline{Y})$ cuya grafica se exhibe en la Figura 4, es unimodal con moda en $\psi = 1.117$ y la masa de probabilidad claramente se acumula en torno a ese valor; en particular $\pi(0.967 < \psi < 1.288 | \underline{Y}) = 0.947$ de

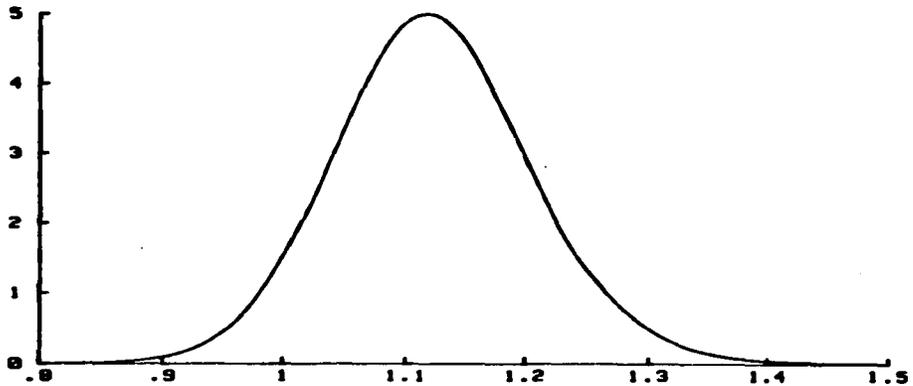


FIGURA 4. Final de referencia $\pi(\psi | \underline{Y})$ datos Tabla 2.

manera que nuevamente los resultados de ambos enfoques resultan coincidentes. En lo que se refiere al parámetro de potencia ρ , aplicando la transformación monótona $\rho = \exp(\psi)$, se tiene que el valor estimado por máxima verosimilitud está dado por $r = 3.069$ con un intervalo de confianza al 95% de $(2.639, 3.626)$. La correspondiente densidad final para ρ se exhibe en la Figura 5 y naturalmente, es también compatible con los resultados frecuentistas.

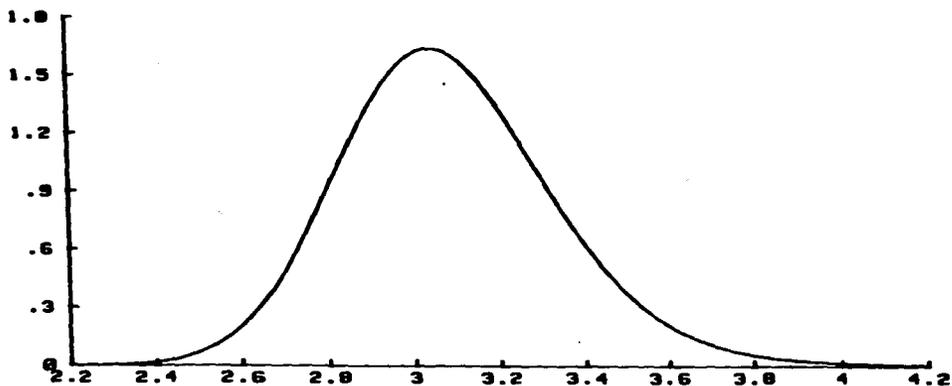


FIGURA 5. Densidad final de referencia $\pi(\rho | \underline{Y})$ datos Tabla 2.

En particular, es unimodal con moda en $\rho = 3.040$ y la masa de probabilidad acumulada alrededor de ese punto. El otro valor crítico de esta densidad se localiza en $\rho = 0.139$ y el comportamiento de la curva en un entorno de ese punto es similar al que se reportó en el ejemplo 1, como se muestra en la Figura 6. La probabilidad $\pi(\rho < 0.2 | \underline{Y})$ es menor que 10^{-8} de manera que este trozo de la densidad puede ignorarse sin mayores

consecuencias prácticas.

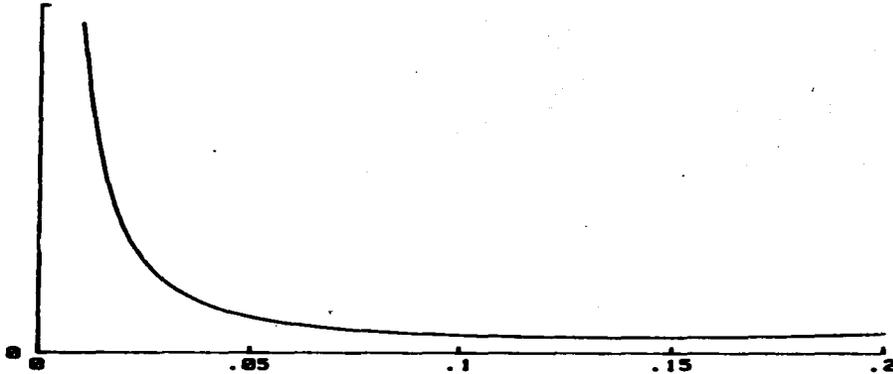


FIGURA 6. La densidad $\pi(p | Y)$ en un entorno de su mínimo local (datos de la Tabla 2).

El siguiente ejemplo también ha sido producido via simulación utilizando la estructura (6.1.3). Presenta una situación en la que el valor del cociente g que se discute en la sección 2.4 es cercano a uno - de hecho, $g = 0.871$ - de manera que puede esperarse un resultado pobre de la región de confianza correspondiente. Los datos se presentan en la Tabla 3 y en terminos de ψ , se tiene que el valor estimado por máxima verosimilitud resulta $f = 0.420$ con un intervalo de confianza de nivel 95% dado por $(-0.358, 11.011)$ cuando el verdadero valor del parametro es $\log(5) = 1.609$. Desde la perspectiva bayesiana, se obtiene la densidad final de referencia $\pi(\psi | \underline{Y})$ que se muestra en la Figura 7. Esta densidad es unimodal con moda en $\psi = 0.781$ y asigna al intervalo $(-0.358, 11.011)$ una probabilidad de 0.96.

TABLA 3. 6

Datos simulados ($n = 5$, $p = q = 4$, $\alpha = 1$, $\beta = 0.6$, $\rho = 5$, $\sigma = 1$)

Estimulo	1				2			
	Dosis 1	2	3	4	0.5	1.5	2.5	3.5
	1.081	2.100	1.192	2.082	0.210	2.284	1.412	2.254
	1.181	2.071	0.919	2.105	2.471	2.421	2.385	2.054
	2.566	2.164	0.620	2.254	2.080	1.096	2.522	2.510
	1.449	0.972	2.050	2.556	1.244	2.672	0.426	2.472
	0.426	1.743	1.120	1.505	1.422	2.126	2.274	0.765

También se puede comprobar que por ejemplo, $\pi(-1.5 < \psi < 4 | \underline{Y}) = 0.954$ y $\pi(-1 < \psi < 3 | \underline{Y}) = 0.925$ de manera que se obtiene información mucho más precisa que la proporcionada por el enfoque frecuentista.

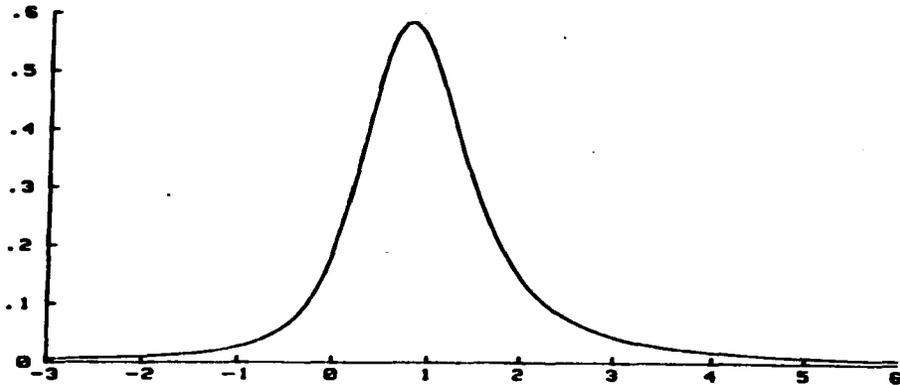


FIGURA 7. Final de referencia $\pi(-\psi | \underline{Y})$ datos Tabla 3.

Si por otra parte, las inferencias se plantean directamente en términos de la potencia ρ , la situación es la siguiente. La estimación por máxima verosimilitud y el correspondiente intervalo de confianza al nivel 95% se obtienen via la transformación respectiva y resultan $r = 2.631$ y $(0.669, 60512.451)$. Mediante el enfoque bayesiano se obtiene la densidad final de referencia $\pi(\rho | \underline{Y})$ cuya gráfica se muestra en la Figura 8.

Nuevamente, la densidad es unimodal con moda en $\rho = 1.712$ y con un mínimo local en $\rho = 0.251$. Es de notar que en este caso la contribución, en términos de masa de probabilidad, del entorno del mínimo local no es despreciable. En cualquier forma, se tiene que $\pi(0.669 < \rho < 60514.451 | \underline{Y}) = 0.96$ como es natural. Pero más aún, se puede verificar que, por ejemplo,

$$\pi(0.251 < \rho < 4.43 | \underline{Y}) = 0.718,$$

$$\pi(0 < \rho < 4.43 | \underline{Y}) = 0.756,$$

$$\pi(0 < \rho < 5 | \underline{Y}) = 0.830,$$

$$\pi(0 < \rho < 10 | \underline{Y}) = 0.90,$$

$$\pi(0.368 < \rho < 20.086 | \underline{Y}) = 0.925$$

y

$$\pi(0.223 < \rho < 54.598 | \underline{Y}) = 0.964$$

de manera que, a pesar de la mayor incertidumbre que existe respecto al parámetro de interés, claramente $\pi(\rho | \underline{Y})$ proporciona información mucho más precisa que la que se puede

obtener con la región de confianza de Fieller, especialmente en términos de la longitud de los intervalos.

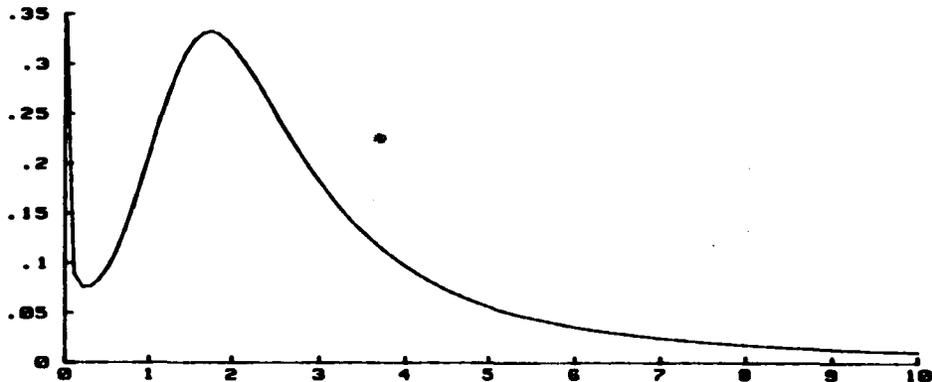


FIGURA 8. Final de referencia $n(\rho | Y)$ datos Tabla 3.

Esta sección concluye con un ejemplo más, también obtenido mediante simulación del modelo (6.1.3). Este es un caso aun más extremo que el anterior en el que la región de confianza de Fieller es absolutamente irrelevante para efectos de inferencia puesto que coincide con toda la recta real. Las condiciones bajo las que aparece este fenómeno fueron discutidas en la sección 2.4. Los datos correspondientes se exhiben en la Tabla 4.

Por lo que respecta a la log-potencia ψ , que tiene un valor de 1.609, los resultados frecuentistas son los siguientes: El valor estimado por máxima verosimilitud es -0.956 y la región de confianza al 95% para ψ coincide con toda la recta real. Si se calcula la densidad final de referencia $n(\psi | Y)$ se obtiene la

TABLA 4.

Datos simulados $Cn = 5$, $p = q = 4$, $\alpha = 2$, $\beta = 0.2$, $\rho = 5$, $\sigma = 1$

Estímulo	1				2			
	Dosis 1	2	4	8	1	3	5	7
	2.081	2.883	1.770	3.266	0.082	2.616	1.561	3.348
	2.181	3.394	1.383	3.089	3.243	2.754	2.713	3.048
	3.566	3.886	1.248	2.938	3.452	1.428	3.651	2.504
	2.450	1.094	2.608	4.140	2.016	3.005	0.564	2.466
	1.436	2.466	1.748	2.149	2.205	2.469	3.502	0.759

gráfica que se muestra en la Figura 9. Como se aprecia ahí, esta densidad es bimodal y sus modas se localizan en $\psi_1 = -2.029$ y en $\psi_2 = 1.776$ con un mínimo local en $\psi = -0.136$.

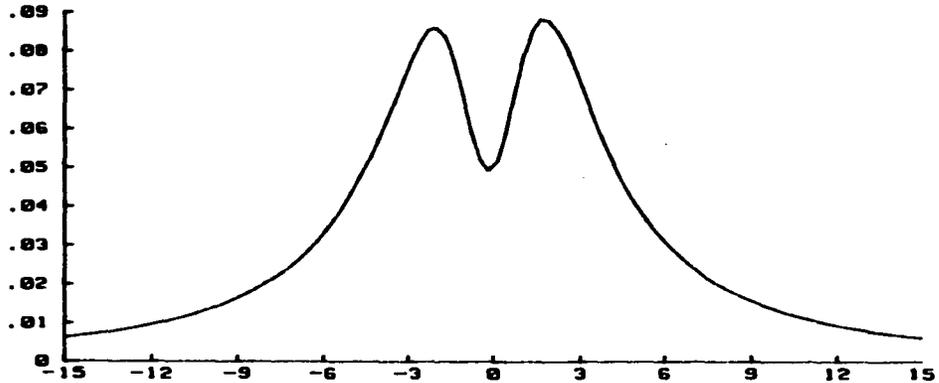


FIGURA 9. Final de referencia $\pi(\psi | \underline{Y})$ datos Tabla 4.

En términos de probabilidad, a partir de esta densidad se puede afirmar que,

$$\pi(0 < \psi < 10 | \underline{Y}) = 0.392.$$

$$\pi(\psi > 0 | \underline{Y}) = 0.499.$$

$$\pi(-10 < \psi < 10 | \underline{Y}) = 0.784.$$

$$\pi(-15 < \psi < 15 | \underline{Y}) = 0.952$$

y

$$\pi(-35 < \psi < 35 | \underline{Y}) = 0.954.$$

Es claro que este es un caso en donde la evidencia experimental proporciona una información modesta sobre el valor del parámetro ψ pero en cualquier forma es también claro que desde el punto de vista bayesiano la descripción de ese estado de incertidumbre relativa no resulta inútil como ocurre con la región de confianza de Fieller.

Cuando las inferencias se realizan directamente con respecto a la potencia ρ , se tiene que el valor del estimador de máxima verosimilitud es $r = 0.384$ y naturalmente, la correspondiente región de confianza al 95% para ρ coincide también, como en el caso de ψ , con toda la recta real. Por otra parte, la final de referencia para ρ se exhibe en la Figura 10. Esta densidad tiene algunas características notables; particularmente, ocurre que esta densidad no tiene puntos críticos, es estrictamente monótona decreciente en el dominio $(0, \infty)$ y en consecuencia no tiene moda alguna. Aún a pesar de esta característica es claro que la masa de probabilidad se acumula

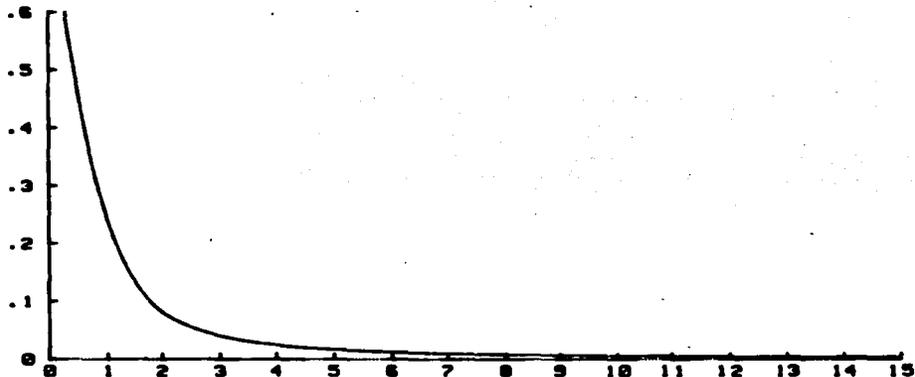


FIGURA 10. Final de referencia $\pi(\rho | \underline{Y})$ datos Tabla 4.

cerca del origen y de hecho, la propiedad de monotonía ocasiona que toda región de máxima densidad sea de la forma $(0, c)$ para algún $c > 0$. En particular se tiene que, por ejemplo,

$$\pi(0 < \rho < 1 | \underline{Y}) = 0.501,$$

$$\pi(0 < \rho < 10 | \underline{Y}) = 0.681,$$

$$\pi(0 < \rho < 20 | \underline{Y}) = 0.734$$

y

$$\pi(0 < \rho < 400 | \underline{Y}) = 0.936$$

de manera que es posible establecer en términos de cuantitativos la muy considerable incertidumbre que existe sobre el parámetro ρ .

CAPITULO 7. CONSIDERACIONES FINALES

Los casos particulares del modelo (1.1.1) que se han utilizado en los capítulos 5 y 6 para ilustrar algunos de los resultados que se pueden obtener si se adopta la propuesta de análisis que se presenta en esta tesis han sido seleccionados debido a que son dos modelos muy comunes en el dominio de los ensayos biológicos; un área que, como ya se indicó en la introducción, da origen de manera natural al problema de estimar un parámetro del tipo cociente. Sin embargo, existen muchos otros modelos, no necesariamente relacionados con aplicaciones biológicas que pueden ser analizados utilizando los resultados de este trabajo. Por ejemplo, el problema de estimar el punto de intersección de dos líneas rectas no paralelas, que tiene relevancia en el estudio de modelos de regresión con puntos de cambio (Smith y Cook 1980), puede ser fácilmente identificado como un caso particular de la estructura analizada.

En cualquier caso, es conveniente reparar en el caso especial del modelo (5.1.1) que se obtiene si $\alpha = 0$. Para este modelo de efectos proporcionales, se tiene que la dimensión del vector de parámetros θ es igual a dos, esto es $k = 2$ y entonces, se aplica la observación que sigue a la expresión (3.3.17). Con las modificaciones oportunas, es posible seguir un desarrollo

similar al de la sección 3.3. para obtener la distribución final del cociente de pendientes que de nuevo resulta el parámetro de interés, esto es, la potencia relativa. Utilizando la notación del capítulo 3 se puede escribir

$$p(\rho | \underline{Y}) \propto p(\rho) [Q(\rho)]^{-1/2} [S_Y^2 - \{S_{XY1} + \rho S_{XY2}\}^2 / Q(\rho)]^{-(N+r-2)/2}$$

en donde

$$N = n(p+q)$$

$$S_Y^2 = \underline{Y}^t \underline{Y} = \sum_j \sum_k (Y_{j,k})^2$$

$$S_{XY1} = \underline{Y}^t \underline{X}_1 = \sum_j \sum_k (Y_{1jk} X_{1j})$$

$$S_{XY2} = \underline{Y}^t \underline{X}_2 = \sum_j \sum_k (Y_{2jk} X_{2j})$$

$$Q(\rho) = c_2 \rho^2 + c_0$$

con

$$c_2 = \underline{X}_2^t \underline{X}_2 = n \sum_j (X_{2j})^2$$

$$c_0 = \underline{X}_1^t \underline{X}_1 = n \sum_j (X_{1j})^2$$

Si se utilizan $p(\rho) = [Q(\rho)]^{-1/2}$ y $r = 2$, entonces se obtiene la distribución final de referencia para ρ

$$p(\rho | \underline{Y}) \propto [Q(\rho)]^{-1} [S_Y^2 - \{S_{XY1} + \rho S_{XY2}\}^2 / Q(\rho)]^{-N/2}$$

que es particularmente interesante porque representa una generalización directa e inmediata de la distribución final de referencia para el cociente de medias Normales. En efecto, basta

con tomar $n = 1$, $X_{1j} = 1$; $j=1, \dots, p$ y $X_{2j} = 1$; $j=1, \dots, q$ para reproducir correctamente los resultados obtenidos por Bernardo (1977).

Otra observación interesante respecto al modelo de cociente de pendientes es la siguiente: Los resultados del capítulo 5 pueden extenderse, utilizando un razonamiento similar al que se ha empleado para el caso de la potencia relativa en el ensayo de líneas paralelas, a una clase más amplia de modelos de este tipo. Si se adopta la suposición de que la respuesta esperada está relacionada linealmente con una función de la dosis, de la forma $f(x) = x^\lambda$, entonces la condición de similaridad garantiza que el parámetro de interés está dado por $\rho^{\lambda/\lambda}$; así, para todos los casos en que $f^{-1}(x)$ está definida en toda la recta real, o bien si se incorpora la hipótesis de que las dosis son positivas, la generalización del procedimiento es inmediata y la distribución final de referencia para la potencia relativa en un ensayo de cociente de pendientes generalizado se obtiene nuevamente en términos de las funciones $Q(\rho)$ y $H(\rho)$.

Volviendo al modelo general, es conveniente recordar que la densidad inicial $p(\rho)$ puede utilizarse para describir el conocimiento inicial - no necesariamente vago o difuso - que posea el científico respecto al parámetro de interés antes de realizar el experimento. Naturalmente, las características analíticas de la correspondiente distribución final dependen de la forma específica en que se decida modelar ese conocimiento y esta circunstancia conduce, también en forma natural, a la consideración de densidades que sean del tipo conjugado. En relación con este tema, en particular, podría utilizarse una densidad inicial de la forma

$$p(\rho) \propto [Q(\rho)]^{-m} \quad ; \quad m > 0$$

que conduce a una final con las mismas propiedades analíticas generales de la final de referencia. Más aún, como un caso límite ($m = 0$) se obtiene otra densidad que podría considerarse no informativa en tanto que es localmente uniforme. Por supuesto que esta familia de densidades 'conjugadas' es muy limitada; por ejemplo, todos sus miembros son unimodales con la misma moda ($\rho = -c_1 / 2c_2$) pero aún así es interesante notar que para los datos del segundo ejemplo de la sección 5.3 las distribuciones finales para todos los elementos de esta familia son prácticamente iguales. En particular, la Figura 1 muestra la densidad final de referencia ($m = 1/2$) y la densidad localmente uniforme ($m = 0$).

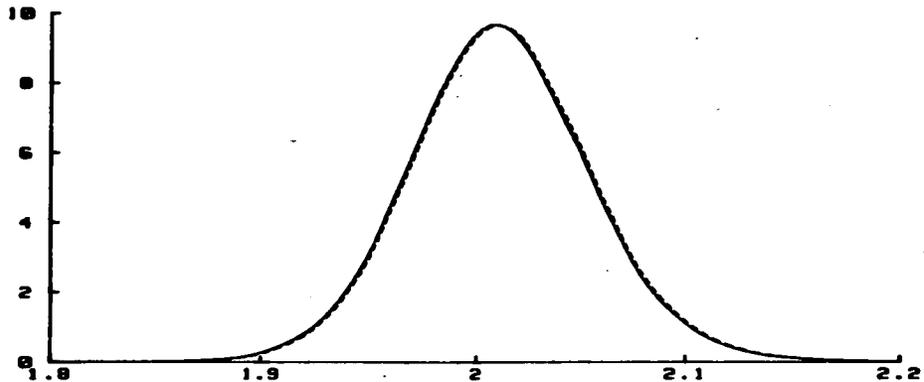


FIGURA 1. Densidades finales ($\pi(c | \underline{y})$: — ; $\rho(c | \underline{y})$: --) para el ejemplo 5.3.2.

En cualquier caso, sería conveniente una investigación más general y en mayor detalle de otras familias de densidades de este tipo pseudo-conjugado que permitiesen incorporar con facilidad

la información inicial.

También a nivel general es oportuno recordar que para todo valor de p estrictamente menor que 1, la correspondiente región de máxima densidad de tamaño p es acotada. Dependiendo de la posible multimodalidad, la región puede estar formada por la unión de dos intervalos o por un sólo intervalo pero es siempre acotada. Esta propiedad resulta de importancia práctica para una interpretación útil de los resultados. Debe observarse, sin embargo que la ventaja de contar con una distribución final es mucho más general puesto que puede utilizarse para establecer la credibilidad final de cualquier región de interés. En particular, y como ejemplo, en modelos de bioensayo de cociente de pendientes, la probabilidad $p(\rho > 0 | Y)$ puede usarse para evaluar la validez del ensayo puesto que un valor negativo de ρ implica que los estímulos tienen un efecto opuesto en la variable de respuesta y la comparación de tal par de estímulos puede considerarse poco afortunada. Más aún, la credibilidad relativa de los intervalos que forman una región de máxima densidad disconexa puede analizarse y algunas distribuciones condicionales, como por ejemplo $p(\rho | Y, \rho > c)$ para alguna constante no-negativa c , pueden calcularse para obtener información más específica sobre ρ . Un análisis cuantitativo completo de este tipo es claramente mucho más complicado desde la perspectiva frecuentista.

Finalmente, existen algunos otros aspectos del modelo y la propuesta de análisis aquí considerada que merecería la pena explorar. Por ejemplo, los resultados centrales de esta tesis no requieren la condición de que la matriz X sea de rango completo puesto que fundamentalmente se necesitan las inversas de las matrices $W^t W$ y $Z_2^t Z_2$. En este sentido, los resultados podrían extenderse para cubrir algunos modelos de rango incompleto. Una conclusión aún más general se obtendría si las matrices inversas

pudiesen reemplazarse por inversas generalizadas. Esta es una línea que queda abierta. De la misma forma, queda abierta la posibilidad de extender los resultados a modelos de MANOVA que en particular resultan de utilidad en el análisis de bioensayos multivariados (Velund 1980, Laska et al. 1985 y Carter y Hubert 1985 por ejemplo) y concretamente en esa área queda abierta la exploración de los procedimientos para la combinación de bioensayos que recientemente ha recibido considerable atención en la literatura (Velund 1982, Srivastava 1986 y Meisner et al. 1986 por ejemplo).

REFERENCIAS

- Anderson, S.A., Bröns, H.K. and Jensen, S.T. (1983). Distribution of eigenvalues in multivariate statistical analysis. *The Annals of Statistics*, 11, 2, 394-415.
- Armitage, P. (1977). *Statistical methods in medical research*. Oxford : Blackwell Scientific Publications.
- Bayes, T.R. (1763). Essay towards solving a problem in the doctrine of chances. Reimpreso en *Biometrika*, 45, (1958), 243-315.
- Berkson, J. (1944). Application of the logistic function to bioassay. *Journal of the American Statistical Association*, 39, 357-365.
- Bermúdez, J.D. (1985). On the asymptotic normality of the posterior distribution in the logistic classification model. *Statistics and Decisions, Supplement issue 2*, 301-308.
- Bernardo J.M. (1977). Inferences about the ratio of normal means : a bayesian approach to the Fieller-Creasy problem. en *Recent Developments in Statistics* (J.R. Barra et al. eds.). New York : North Holland.
- Bernardo J.M. (1979). Reference posterior distributions for bayesian inference (con discusion). *Journal of the Royal Statistical Society, B*, 41, 113-147.
- Bliss, C.I. (1935a). The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22, 134-167.
- Bliss, C.I. (1935b). The comparison of dosage-mortality data. *Annals of Applied Biology*, 22, 307-333.
- Bliss, C.I. (1952). *The statistics of bioassay*. New York : Academic Press.
- Bliss, C.I. (1970). *Statistics in biology*, vol.2. New York : McGraw-Hill.
- Bon, G.E.P. and Tiao, G.C. (1973). *Bayesian inference in statistical analysis*. Reading, Mass. : Addison-Wesley.

- Buonaccorsi, J.P. (1986). Designs for slope ratio assays. *Biometrics*, 42, 875-882.
- Carter, E.M. and Hubert, J.J. (1985). Analysis of parallel-line assays with multivariate responses. *Biometrics*, 41, 703-710.
- Chen, C. (1985). On asymptotic normality of limiting density functions with bayesian implications. *Journal of the Royal Statistical Society, B*, 47, 540-546.
- Creasy, M.A. (1954). Limits for the ratio of normal means. *Journal of the Royal Statistical Society, B*, 16, 186-194.
- Darby, S.C. (1980). A bayesian approach to parallel line bioassay. *Biometrika*, 67, 3, 607-612.
- Dawid, A.P. (1970). On the limiting normality of posterior distributions. *Proceedings of the Cambridge Philosophical Society*, 67, 625-633.
- DeFinetti, B. (1974). Bayesianism : its unifying role for both the foundations and applications of statistics. *International Statistical Review*, 42, 2, 117-130.
- DeGroot, M.H. (1970). *Optimal statistical decisions*. McGraw-Hill : New York.
- Dickey, J.M. (1976). Approximate posterior distributions. *Journal of the American Statistical Association*, 71, 355, 680-690.
- Dickey, J.M. (1973). Scientific reporting and personal probabilities : Student hypothesis. *Journal of the Royal Statistical Society, B*, 35, 285-305.
- Edwards, W., Lindman, H. and Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242. Reimpresso in *The writings of Leonard Jimmie Savage - A memorial selection*. (1981). Washington : The American Statistical Association and The Institute of Mathematical Statistics.
- Emery, W.B., Lees, K.A. and Tootill, J.P.R. (1951). The assay of vitamin B12, part IV. The microbiological estimation with *Lactobacillus leichmannii* 113 by the turbidimetric procedure. *Analyst*, 76, 141-146.
- Fieller, E.C. (1940). The biological standarization of insulin. *Journal of the Royal Statistical Society, Supplement*, 7, 1-84.

- Fieller, E.C. (1944). A fundamental formula in the statistics of biological assay and some applications. *Quarterly Journal of Pharmacy and Pharmacology*, 17, 2, 117-123.
- Fieller, E.C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society, B*, 16, 175-185.
- Finney, D.J. (1952/78). *Statistical method in biological assay*. 3rd ed.. London : Griffin.
- Fraser, D.A.S. and McDunnough, P. (1984). Further remarks on asymptotic normality of likelihood and conditional analysis. *The Canadian Journal of Statistics*, 12, 3, 183-190.
- Gradshteyn, I.S. and Ryzhik, I.M. (1985). *Tables of integrals, series and products*. New York : Academic Press.
- Healy, M.J.R. (1949). Routine computation of biological assays involving a quantitative response. *Biometrics*, 5, 330-334.
- Heyde, C.C. and Johnstone, I.M. (1979). On asymptotic posterior normality for stochastic processes. *Journal of the Royal Statistical Society, B*, 41, 184-189.
- Irwin, J.O. (1937). Statistical method applied to biological assay. *Journal of the Royal Statistical Society, Supplement*, 4, 1-60.
- Johnson, R.A. (1967). An asymptotic expansion for posterior distributions. *The Annals of Mathematical Statistics*, 38, 1899-1907.
- Johnson, R.A. (1970). Asymptotic expansions associated with posterior distributions. *The Annals of Mathematical Statistics*, 41, 851-864.
- Kappenman, R.F., Geisser, S. and Antle, C.E. (1970). Bayesian and fiducial solutions for the Fieller-Creasy problem. *Sankhya, B*, 32, 331-340.
- Kendall, M.G. and Stuart, A. (1979). *The advanced theory of statistics*, vol. 2. 4th. ed.. London : Griffin.
- Knudsen, L.F. and Curtis, J.M. (1947). The use of the angular transformation in biological assays. *Journal of the American Statistical Association*, 42, 282-296.
- Laska, E.M. Kushner, H.B. and Meisner, M. (1985). Multivariate bioassay. *Biometrics*, 41, 547-554.

- LeCam. L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics*, 1, 277-330.
- LeCam. L. (1956). On the asymptotic theory of estimation and testing hypothesis. *Proceedings of the 3rd Symposium on Mathematical Statistics and Probability*, vol. 1, 129-156. Berkeley : University of California Press.
- Lindley. D.V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27, 986-1005.
- Lindley. D.V. (1971). *Bayesian statistics, a review*. Regional Conference. Series on Applied Mathematics. SIAM.
- Meisner. M., Kushner. H.B. and Laska. E.M. (1966). Combining multivariate bioassays. *Biometrics*, 42, 421-427.
- Mendoza. M. (1987). A bayesian analysis of the slope ratio bioassay. *Enviado para publicación*.
- Morrison. D.F. (1976). *Multivariate statistical methods*. McGraw-Hill : Tokyo.
- Rao. C.R. (1973). *Linear statistical inference and its applications*. 2nd. ed.. New York : Wiley.
- Savage. L.J. (1977). The shifting foundations of statistics. in *Logic, laws and life : some philosophical complications*. (Colodny, R. ed.) Pittsburgh : University of Pittsburgh Press, 3-18. Reimpreso in *The writings of Leonard Jimmie Savage - A memorial selection*. (1981). Washington : The American Statistical Association and The Institute of Mathematical Statistics.
- Scheffe. H. (1959). *The analysis of variance*. New York : Wiley.
- Searle. S.R. (1971). *Linear models*. New York : Wiley.
- Searle. S.R. (1982). *Matrix algebra useful for statistics*. New York : Wiley.
- Serfling. R.J. (1980). *Approximation theorems of mathematical statistics*. New York : Wiley.
- Shannon. C.E. (1948). A mathematical theory of communications. *Bell System Technical Journal*, 27, 379-423, 623-656.

- Smith. A.F.M. and Cook. D.G. (1980). Straight lines with change-point : a bayesian analysis of some renal transplant data. *Applied Statistics*. 29. 180-189.
- Srivastava. M.S. (1986). Multivariate bioassay, combination of bioassays and Fieller's theorem. *Biometrics*. 42, 131-141.
- Tsutakawa. R.K. (1982). Statistical methods in bioassay. in *Encyclopedia of statistical sciences*. vol. 1. (Kotz. S. and Jhonson. N.L. eds.). 236-243.
- Volund. A. (1982). Combinations of multivariate bioassay results. *Biometrics*. 38. 181-190.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*. 20. 595-601.
- Walker. A.M. (1969) On the asymptotic behaviour of posterior distributions. *Journal of the Royal Statistics Society B* 31. 80-88.
- Williams. D.A. (1986). Interval estimation of the median lethal dose. *Biometrics*. 42. 641-645.
- Wood. E.C. (1946). The theory of certain analytical procedures, with particular reference to micro-biological assays. *Analyst*. 71. 1-14.
- Zerbe. G.O. (1978). On Fieller's theorem and the general linear model. *The American Statistician*. 32, 103-105.