

UNIVERSIDAD DE LA LAGUNA

**«Heurísticas bayesianas para el
problema de control multiperiodo»**

**Autor: Enrique Francisco González Dávila
Director: Dr. D. Josep Ginebra Molins**

**Departamento de Estadística e
Investigación Operativa y Computación**

UNIVERSIDAD DE LA LAGUNA
FACULTAD DE CIENCIAS MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA, INVESTIGACIÓN OPERATIVA Y
COMPUTACIÓN

HEURISTICAS BAYESIANAS PARA EL PROBLEMA DE CONTROL
MULTIPERIODO

Enrique Fco. González Dávila.

Memoria para optar al grado de Doctor en
Ciencias Matemáticas realizada bajo la
dirección del Dr. D. Josep Ginebra Molins.

La Laguna, 1999

A mis padres, a Norena y a mis
hermanos José Carmelo, Melchor y Cande.

Este trabajo ha sido realizado bajo la dirección del Dr. D. Josep Ginebra Molins, a quien quiero expresar mi más sincero agradecimiento, por la dedicación e inestimable ayuda prestada a lo largo de todo el periodo de su realización, salvando incluso las dificultades de la distancia.

También quiero hacer constar mi agradecimiento a los compañeros del Departamento, en particular al Dr. D. Miguel Ángel González Sierra, sin el cual nunca hubiera sido posible realizar este trabajo, y a los compañeros de la segunda planta de la Torre de Química por su aliento y ayuda.

CONTENIDOS

Parte I: El problema de control multiperiodo.	1
Capítulo 1: Introducción.	3
1.1 Descripción del problema de control multiperiodo.	3
1.2 Compromiso entre aprendizaje y control.	6
1.3 Repaso bibliográfico: Diseños de Experimentos.	7
1.4 Modelo lineal.	10
1.5 Diseños óptimos.	13
1.6 Descripción de los contenidos.	15
1.7 Notación.	19
Capítulo 2: Diseños <i>certainty equivalence</i> y diseños <i>miopes</i>.	21
2.1 Diseños <i>certainty equivalence</i> .	22
2.1.1 Repaso bibliográfico.	23
2.1.2 Implementación para modelo lineal.	25
2.1.3 Diseño <i>certainty equivalence</i> no secuencial.	26
2.2 Comparación de <i>certainty equivalence</i> secuencial con no secuencial.	27
2.3 Diseños <i>miopes</i> .	29
2.4 Comparación de <i>certainty equivalence</i> frente a <i>miope</i> .	31
2.5 Necesidad de perturbar <i>miope</i> y <i>certainty equivalence</i> .	32
Apéndice 2.A Demostración Proposición 2.1.	34
Parte II: Una sola variable de control.	37
Capítulo 3: Diseños <i>Fieller</i>.	39
3.1 Los intervalos <i>Fieller</i> .	40
3.2 Los diseños <i>Fieller</i> .	42
3.2.1 Definición de la familia <i>Fieller</i> .	42
3.2.2 Búsqueda del mejor diseño <i>Fieller</i> .	46
3.3 Diseño <i>Fieller</i> para modelo lineal.	47
3.4 Análisis de sensibilidad.	50

3.5	Extensiones y comentarios.	53
	Apéndice 3.A Cálculo de intervalos <i>Fieller</i> para el modelo lineal.	55
3.A.1	Priori conjugada con \mathbf{b}_0 independiente de \mathbf{b}_1 .	55
3.A.2	Priori no conjugada.	59
Capítulo 4:	Diseños basados en regiones creíbles.	63
4.1	Regiones creíbles.	64
4.1.1	Regiones <i>HPD</i> .	65
4.1.2	Intervalos centrales a posteriori.	66
4.2	Diseños para control multiperiodo basados en regiones <i>HPD</i> .	66
4.3	Aplicación de los diseños <i>HPD</i> al modelo lineal.	68
4.4	Diseños para el <i>RSB</i> basados en intervalos centrales a posteriori.	71
4.5	Aplicación de los diseños “ <i>CP</i> ” al modelo lineal.	73
4.6	Extensiones y comentarios.	74
	Apéndice 4.A Marginal de \mathbf{q} para el ejemplo de la Sección 4.3.	77
Capítulo 5:	Diseños que adaptan las secuencias de aproximación estocástica	
	(Diseños <i>san</i>).	79
5.1	Secuencias de aproximación estocásticas.	80
5.2	Diseños basados en la modificación de las aproximaciones estocásticas.	82
5.2.1	Definición de diseños <i>san</i> .	82
5.2.2	Interpretación de diseños <i>san</i> .	83
5.2.3	Búsqueda del mejor diseño <i>san</i> .	87
5.3	Aplicación al modelo lineal.	88
5.4	Análisis de sensibilidad.	90
5.5	Extensiones y comentarios.	93
	Apéndice 5.A Demostración Proposición 5.1.	95
Parte III:	Múltiples variables de control.	99
Capítulo 6:	Diseños <i>híbridos</i>.	101
6.1	Definición de diseños <i>híbridos</i> .	102
6.2	Diseños <i>híbridos</i> para el modelo lineal.	105
6.2.1	Modelo lineal de primer orden.	105
6.2.2	Modelo lineal con priori conjugada (\mathbf{s}^2 conocida).	108

6.2.3	Modelo lineal con priori conjugada (\mathbf{s}^2 desconocida).	110
6.2.4	Modelo lineal de segundo orden.	111
6.3	Aplicación al modelo lineal.	112
6.3.1	Ejemplo para modelo lineal simple ($k = 1$).	113
6.3.2	Ejemplo para modelo lineal múltiple ($k = 3$).	114
6.4	Análisis de sensibilidad.	115
6.4.1	Ejemplo para modelo lineal simple ($k = 1$).	116
6.4.2	Ejemplo para modelo lineal múltiple ($k = 3$).	118
6.5	Extensiones y comentarios.	120
	Apéndice 6.A.1 Varianza de predicción.	122
	Apéndice 6.A.2 Demostración Proposición 6.1 y 6.3.	123
	Apéndice 6.A.3 Demostración Lema 6.2 y Proposición 6.2.	127
	Apéndice 6.A.4 Cálculo de $x^{hb}(H)$ para un modelo lineal bajo la restricción de que $C = B(R)$.	129
	Apéndice I. Resultados distribucionales: Modelo de regresión simple normal con priori sobre \mathbf{b}_0 y \mathbf{b}_1 .	133
	Apéndice II. Resultados distribucionales: Modelo de regresión simple normal parametrizado en términos de \mathbf{b}_1 y \mathbf{q} .	151
	 Bibliografía.	 165

Parte I:

El problema de control multiperiodo

CAPITULO 1.

INTRODUCCION

1.1 Descripción del problema de control multiperiodo.

Asumamos que podemos controlar una entrada $x_n = (x_{n1}, \dots, x_{nk}) \in C \subset R^k$ y observar una variable respuesta y_n tal que:

$$y_n(x_n) = f(x_n; \mathbf{b}) + \mathbf{e}_n, \quad n = 1, \dots, N$$

con $E[y|x, \mathbf{b}] = f(x; \mathbf{b})$ una función continua, \mathbf{b} el vector de parámetros sobre los que se dispone una distribución a priori, $\mathbf{p}_0(\mathbf{b})$, y \mathbf{e}_n el error con que se observa y_n . En los ejemplos supondremos que las \mathbf{e}_n son independientes e idénticamente distribuidas según una normal con varianza constante, $Var[\mathbf{e}_n | x, \mathbf{b}] = \mathbf{s}^2$. Sin embargo, la metodología presentada es aplicable a otras distribuciones y consideraciones sobre la varianza, como puede ser el caso en el que la distribución a priori es conjunta sobre \mathbf{b} y sobre \mathbf{s}^2 , tal como se muestra en el capítulo 6. Además se considerará que la región experimental C es compacta.

En el problema de control multiperiodo se desea elegir secuencialmente valores para las variables de entrada, x_n , de forma que la respuesta y_n esté siempre cerca de un valor objetivo conocido T . Como soluciones al mismo, en principio se considerarán diseños totalmente secuenciales, denotados por d , que seleccionan en cada estadio el punto del diseño

$x_{n+1}^d \in C \subset R^k$ dadas las anteriores entradas, $X_n^d = (x_1^d, \dots, x_n^d)$, y observaciones, $Y_n' = (y_1(x_1^d), \dots, y_n(x_n^d))$. El objetivo es encontrar diseños secuenciales que minimicen $r_i(d) = E[L_i(X_N^d, Y_N, T)]$, donde $L_i(X_N^d, Y_N, T)$ es una función de pérdida que penaliza la distancia entre Y_N' y T , y donde la esperanza es respecto a la distribución conjunta en el estadio cero de los parámetros y de la respuesta y . Notar que a diferencia de las funciones de pérdida clásicas, que sólo dependen del parámetro y de la decisión, en los problemas de control estocástico la función de pérdida además depende del valor observado y .

Como casos particulares se consideran la minimización de:

$$r_1(d) = E[\sum_{n=1}^N |y_n(x_n^d) - T|] = E[L_1(X_N^d, Y_N, T)],$$

$$r_2(d) = E[\sum_{n=1}^N (y_n(x_n^d) - T)^2] = E[L_2(X_N^d, Y_N, T)],$$

y el problema de dos etapas

$$r_3(d) = E[\sum_{n=1}^{N'} (y_n(x_n^d) - T)^2 + S(y_{N'+1}(x_{N'+1}^d) - T)^2] = E[L_3(X_N^d, Y_N, T)]$$

con $S > 0$ conocido. Esta última función de pérdida podrá modelar la situación planteada en muchos experimentos clínicos donde las primeras N' observaciones corresponden a la etapa experimental y S al peso del valor de la decisión final adoptada. Como límite de L_3 , cuando S es mucho más grande que N , aparece el problema de estimación de la solución de $f(x; \mathbf{b}) = T$, ya que la última observación es la que tendrá todo el peso, aunque en esta memoria no se explora esta situación. Generalizaciones inmediatas de estas pérdidas se pueden obtener ponderando la aportación de cada observación a la pérdida global a través de un valor \mathbf{a}_n conocido, como podría ser:

$$r_1^g(d) = \sum_{n=1}^N \mathbf{a}_n |y_n - T| \quad \text{ó} \quad r_2^g(d) = \sum_{n=1}^N \mathbf{a}_n (y_n - T)^2.$$

En los casos en los que se quiera penalizar de manera distinta el observar por encima o por debajo del valor objetivo T se podrían utilizar pérdidas asimétricas, tal como:

$$r_1^{as}(d) = \sum_{n=1}^N \mathbf{a}_n |y_n - T| + \mathbf{x}_n (y_n - T).$$

Existen muchas situaciones en las que la respuesta observada y_n en el punto x_n no será conocida a la hora de elegir el siguiente punto a observar x_{n+1} . A pesar que dicho caso no se trata en esta tesis, lo que proponemos se podría adaptar fácilmente a dicho caso a base de restringir las soluciones a diseños solo parcialmente secuenciales.

Son muchos los ejemplos prácticos donde se deben resolver problemas de control multiperiodo. Por ejemplo, en Economía, Ingeniería o en Experimentos Clínicos es corriente encontrar problemas en los que conviene elegir los sucesivos niveles de las variables de entrada de forma que las respuestas y_1, \dots, y_n estén próximas a un valor objetivo conocido T , a la par que interesa descubrir el valor de la variable x en la que $E[y|x] = T$ y, por tanto, estimar la raíz \mathbf{q} de $f(x; \mathbf{b}) = T$.

Así, problemas de búsqueda secuencial de dosis óptima, dosis letal (“ LD_p ”) y estimación de percentiles en curvas de respuesta cuantal, en los que no sólo interese la estimación de la raíz sino que donde la distancia $|y - T|$ tiene un significado en términos de daño potencial al enfermo, son ejemplos típicos de aplicación. Para más detalle sobre este tipo de problemas ver McLeish y Tosh (1983, 90), Wu(1988), Efron (1971), Robbins y Siegmund (1974) y Zacks y Eichhorn (1975). Otras aplicaciones en los que se pueda dar la misma interpretación anterior son los problemas de calibración, resumidos en Buonarccorsi (1986), y los problemas que intentan identificar condiciones que minimizan la variabilidad de algunas características de calidad, como la razón señal ruido, estudiados por ejemplo en Box (1988).

1.2 Compromiso entre aprendizaje y control.

Cuando se selecciona x_n para los problemas de control multiperiodo nos enfrentamos a un conflicto difícil de resolver. Por un lado, se necesita elegir x_n de forma que la distancia entre el valor observado $y_n(x_n)$ y T sea pequeño en valor esperado, pero por otro lado, se va a usar $(x_n, y_n(x_n))$ para aprender sobre $f(x; \mathbf{b})$ y ayudar a mejorar el control en las fases posteriores. Tal como se muestra en la Figura 1.1, cada uno de estos dos objetivos por separado obligarían a situar a x_n en partes muy diferentes de la región experimental C . Si lo único que importa es aprender sobre la superficie $f(x; \mathbf{b})$, o sobre una característica de la misma, como puede ser en el caso de $k=1$ la raíz de $f(x; \mathbf{b})=T$, el punto x_n deberá situarse cercano a la frontera de la región C , de manera que los puntos x_i ($i=1, \dots, n$) queden muy dispersos tal como dicta la teoría clásica de diseños óptimos. Si en cambio lo que se quiere es elegir el x_n que minimice una distancia en valor esperado entre $y_n(x_n)$ y T , se deberá situar el punto x_n cerca del valor previsto para la raíz de $f(x; \mathbf{b})=T$ en el estadio n .

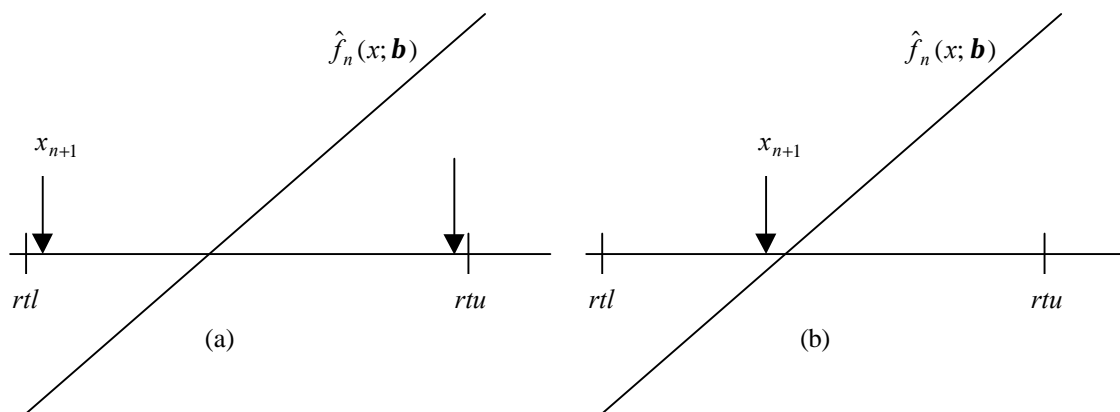


Figura 1.1 Cuando se quiere aprender sobre \mathbf{b} , lo conveniente es trabajar con diseños en los que se experimenta en puntos muy dispersos de la región de experimentación, concentrados todos cerca de las fronteras de la misma, como es el caso (a). En cambio, para control, como refleja el gráfico (b), para una buena elección de x_{n+1} sólo importa minimizar en promedio la diferencia entre $y_{n+1}(x_{n+1})$ y T .

Ya en Prescott (1972) se hace referencia a este dilema existente entre la necesidad de aprender sobre la superficie y la necesidad de evitar una pérdida inmediata excesiva para el problema de control cuando $x \in R$ y $f(x; \mathbf{b}) = x\mathbf{b}$. Este dilema es lo que en la literatura de control adaptativo se denomina el aspecto dual del control, tal y como desarrolla el Capítulo 7 de Astrom y Wittenmark (1995).

En el caso de un único control, se asume que se puede controlar un input $x_n \in C \subset R$, con C compacto y que existe una única raíz $\mathbf{q} \in R$ tal que $f(\mathbf{q}; \mathbf{b}) = T$. En estas circunstancias, el modelo lineal suele ser una buena aproximación a $f(x; \mathbf{b})$ en un entorno de $x = \mathbf{q}$. Notar además que por ser T conocido, y sin pérdida de generalidad, el problema siempre se puede reducir al caso $T = 0$ a base de trabajar con $y'_n(x_n) = y_n(x_n) - T$ de forma que $y'_n(x_n) = y_n(x_n) - T = (f(x_n; \mathbf{b}) - T) + \mathbf{e}_n$. Dado que cuando se tiene más de un control, y por tanto $k > 1$, la solución de $f(x; \mathbf{b}) = E[y|x, \mathbf{b}] = T$ no será un valor real sino una hipersuperficie, se complica el problema y nos obliga a tratar por separado los casos $k = 1$ y $k > 1$. Aunque existe gran cantidad de referencias sobre el problema de control multiperiodo, la gran mayoría de ellas se limitan a estudiar el caso de un único control, $k = 1$.

1.3 Repaso bibliográfico: Diseños de Experimentos.

En esta sección se realiza un repaso bibliográfico sobre los criterios y objetivos perseguidos por la teoría de diseños de experimentos en diferentes problemas. La meta que se persigue es la de clarificar y exponer la metodología, así como el fin por el que se lleva a cabo el diseño de experimentos en distintos problemas, tal como puede ser en la teoría clásica de diseños óptimos, y con ello comparar la metodología y el fin con que se realizan los diseños para problemas de control multiperiodo. Esto servirá para poder contrastar con el repaso

bibliográfico que se realiza en la Sección 2.1.1 de esta memoria sobre los diseños más utilizados en problemas de control multiperiodo, los diseños *certainty equivalence*.

En la teoría de diseños óptimos, el objetivo es aprender sobre $f(x; \mathbf{b})$ o sobre una característica de $f(x; \mathbf{b})$ tan eficientemente como sea posible. En este contexto los diseños son realizados con el objetivo de aprender sobre los parámetros del modelo, y una manera de medir esta eficiencia es en términos de la varianza de los parámetros estimados del modelo. Así, se pueden construir diseños que observan en puntos que minimizan el determinante o la traza de la varianza de los parámetros estimados (diseños D-óptimos o A-óptimos, respectivamente) o que maximizan la varianza de predicción (diseños G-óptimos). Si se asume que el modelo es lineal en los parámetros, $E[Y] = \mathbf{Z}\mathbf{b}$, esto se traslada en criterios de optimalidad basados sobre funciones de $\mathbf{Z}'\mathbf{Z}$ muy relacionados con la matriz de información de Fisher de \mathbf{b} . Buenas referencias de este campo son Fedorov (1972) y Silvey (1980). Es importante destacar que en este tipo de diseños los puntos elegidos suelen quedar repartidos a lo largo de la región experimental, con frecuencia en la frontera, y que no tienen en cuenta para nada los valores de la respuesta observada en dichos puntos.

Atkinson y Donev (1992), Chaloner (1989), Frees y Ruppert (1990) y Pukelsheim (1993) revisan de forma bastante completa los diseños óptimos no Bayesianos para modelos lineales y Ford, Kitsos y Titterington (1989) hacen lo propio para modelos no lineales. Chaloner y Verdinelli (1996) resumen los resultados sobre diseños óptimos Bayesianos tanto para modelos lineales como no lineales.

En cuanto a diseños específicos para estimar la raíz \mathbf{q} , Wu (1988) busca secuencialmente la dosis óptima y estima percentiles en respuesta cuantal, Buonaccorsi (1986) y otras referencias allí citadas estudian diseños para calibración, y Box (1988), Chaloner (1989), Frees y Ruppert (1990) y Buonaccorsi y Iyer (1986) obtienen diseños que minimizan la variabilidad de características de control.

En un contexto diferente, la teoría de aproximación estocástica, que parte de Robbins y Monro (1951) y de Kiefer y Wolfowitz (1959), estudia el problema de estimar las raíces u óptimos de funciones que únicamente pueden ser observadas con ruido. Ellos dan la solución en términos de una secuencia recursiva de variables aleatorias x_n en el espacio de las variables de entrada que converge casi seguro al máximo o a la raíz de la función. Estas secuencias son definidas con el propósito de obtener normalidad asintótica y convergencia en probabilidad casi segura hacia el óptimo y la raíz. Lai y Robbins (1979, 82), Wu (1986), Ruppert (1991), Ginebra (1996) y otros muchos, utilizan modificaciones y casos particulares de estas secuencias.

Los problemas de estos dos últimos párrafos se distinguen de los problemas de control multiperiodo en que en los de control se intenta que la distancia en promedio entre las observaciones y un valor objetivo T sea mínima y, por tanto, que se realice el mayor número de observaciones sobre la raíz de $f(x; \mathbf{b}) = T$. En los casos anteriores siempre se podría intentar observar en puntos muy alejados de la raíz, aunque se empeore el control, con el fin de aumentar la convergencia final del diseño hacia esa raíz.

La optimización de funciones con ruido, desde el punto de vista del control multiperiodo, requieren un compromiso entre el aprendizaje sobre la superficie y el observar tantos experimentos como sea posible en el máximo o en la raíz. En el control de procesos estocásticos y en las pruebas clínicas existen costos e imperativos éticos que exigen seleccionar los mejores valores para las variables de entrada para el individuo que debe ser tratado a continuación, a la vez que se debe aprender sobre la superficie tal como describen Astrom y Wittenmark (1995) y Rosenberger (1996). Ginebra y Clayton (1995) definen el *Response Surface Bandit (RSB)* para resolver problemas donde los diseños deben ser llevados a cabo por el costo en términos de respuesta observada y no por el conocimiento del modelo, enfocando la atención en la maximización de la utilidad esperada que mide la cercanía entre Y'_N y el óptimo de $f(x; \mathbf{b})$.

En general, las decisiones realizadas bajo incertidumbre requieren elegir una acción d que maximice una utilidad esperada, la función de los parámetros y de las observaciones. Spiegelhalter et al. (1994), Berry (1993) y Berry y Stangl (1996) discuten aspectos generales relacionados con el uso de métodos de diseños óptimos Bayesianos en problemas de decisión en Medicina. En muchas aplicaciones las decisiones son ejecutadas secuencialmente, y así la política a seguir depende de las anteriores decisiones y observaciones realizadas. Tales problemas de decisión son notoriamente difíciles de resolver. En general, el material sobre análisis secuencial Bayes, ver por ejemplo Berger (1985), suele estar muy relacionado con otras dos áreas extensivamente estudiadas: la programación dinámica y los problemas de parada óptima. El enfoque del problema tratado en esta memoria no se puede catalogar como de un problema de parada óptima, puesto que se ha considerado que el número N de pruebas a realizar está fijo por el experimentador y que, en general, no se tratará de un número excesivamente grande. Para ejemplos de resolución de problemas de parada óptima ver Müller et al. (1999).

1.4 Modelo lineal.

En principio, la función $f(x; \mathbf{b})$ podría pertenecer a una familia cualquiera de funciones paramétricas continuas pero, para problemas con una sola variable de control, esta función cerca de la raíz $x = \mathbf{q}$ se podrá aproximar casi siempre por un modelo de regresión simple. Estudiando el problema de control multiperiodo suponiendo que la distribución de y conocidos x y \mathbf{b} , sigue un modelo lineal se resolverá el caso en el que dicho modelo sea el más realista y se obtendrá un buen diseño para cuando el modelo lineal sea una buena aproximación al modelo real.

Todos los métodos que se propondrán serán métodos Bayesianos y como tales necesitan de la densidad de probabilidad condicional de y dado $x = (x_1, \dots, x_k)'$, $g(y | \mathbf{b}, x)$, y de una distribución a priori sobre los parámetros \mathbf{b} y \mathbf{s}^2 que recoja la información que se dispone sobre los mismos. Al analizar datos puede plantearse el debate sobre si tiene sentido o no incorporar información a priori sobre los parámetros, sin embargo, en el contexto del diseño de experimentos este debate no tiene sentido ya que al no haberse recogido aún los datos, los observables y los parámetros están a un mismo nivel. Es lógico intentar aprovechar al máximo toda la información de que se disponga para escoger los puntos de experimentación con todas las garantías. Esto es especialmente relevante al diseñar experimentos para un número pequeño de observaciones.

Sea $Y_n = (y_1, y_2, \dots, y_n)'$ el vector de las n primeras respuestas y_i y Z_n la matriz $n \times (k+1)$ donde la i -ésima fila son los valores x_i para los que se ha observado y_i . Para el modelo lineal se tiene que:

$$E[y_i | \mathbf{b}, x_i] = \mathbf{b}_0 x_{i0} + \mathbf{b}_1 x_{i1} + \dots + \mathbf{b}_k x_{ik}.$$

A menudo la componente x_{i0} es fijada a 1 para todo i , de manera que $\mathbf{b}_1 x_{i0}$ es constante y Z_n se puede escribir como $Z_n = (\mathbf{1}_n : X_n)$. En cuanto a la parte estocástica de estos modelos la versión más utilizada es el modelo normal, en el cual la distribución de Y_n dado X_n es una normal.

Los ejemplos de esta memoria se restringen al modelo lineal normal en el que las varianzas condicionales son iguales para todo i , $Var[y_i | \mathbf{b}, x_i] = \mathbf{s}^2$, y en el que las observaciones y_i son condicionalmente independientes. A la hora de la implementación de estos ejemplos se ha considerado el caso particular en el que \mathbf{s}^2 es conocida, y aunque pueda parecer una restricción muy fuerte, se conseguirá reducir el tiempo de computación ya de por sí muy elevado para algunas de las propuestas introducidas. Cuando los resultados obtenidos sean válidos para el caso en que sobre \mathbf{s}^2 conozca una distribución a priori se indicará. En el

capítulo 6 se ejemplifica una situación en la que la distribución a priori es tomada sobre $(\mathbf{b}, \mathbf{s}^2)$, con $1/\mathbf{s}^2$ siguiendo una Gamma y \mathbf{b} una t-student multivariante a priori.

La regresión lineal se introduce en muchos libros como, por ejemplo, Weisberg (1985) y Neter, Wasserman y Kutner (1989). Buenas referencias sobre la regresión lineal Bayesiana son los clásicos Zellner (1971) y Box y Tiao (1973) y más recientemente Gelman et al (1995). Son abundantes los libros donde se describen métodos computacionales para regresión lineal, por ejemplo, Gill, Murray y Wright (1981) y Golub y Van Loan (1983). El análisis de residuos en regresión en un contexto Bayesiano aparece en Zellner (1975), Chaloner y Brant (1988) y Chaloner (1991). Boscardin y Gelman (1996) presentan una variedad de modelos paramétricos para varianzas no constantes desde el punto de vista Bayesiano.

Una vez especificada la distribución a priori sobre \mathbf{b} o sobre $(\mathbf{b}, \mathbf{s}^2)$, las propuestas introducidas para el problema de control multiperiodo con una sola variable de control requieren el cálculo de la distribución a posteriori de la raíz, del valor previsto de y en x , de una estimación de su desviación típica y de la pendiente de $E[y|x, \mathbf{b}]$.

Para facilitar la implementación de los mismos, en los Apéndices I y II se recogerán algunos ingredientes necesarios a la hora de calcular todos los valores listados en el párrafo anterior para ocho prioris distintas sobre \mathbf{b} . Las cuatro primeras contemplan las cuatro posibles combinaciones de las distribuciones uniforme y normal sobre los parámetros \mathbf{b}_0 y \mathbf{b}_1 . En el Apéndice II se reparametriza $E[y|x, \mathbf{b}]$ a través de $E[y|x, \mathbf{b}] = T + \mathbf{b}_1(x - \mathbf{q})$ y se asume que las prioris sobre $(\mathbf{b}_1, \mathbf{q})$ son una de las cuatro combinaciones de las distribuciones uniforme y normal sobre ellos.

Para las implementaciones en el modelo con más de una variable de control sólo se necesitan los dos primeros momentos de la distribución a posteriori de \mathbf{b} . En el Apéndice I se presentan los mismos para el caso de priori conjugada. Para otras prioris más complejas habrá

que obtener dichos momentos generalmente utilizando técnicas de simulación de la posteriori, tales como MonteCarlo y muestreo Gibbs.

1.5 Diseños óptimos.

El hecho de que gran cantidad de autores sigan estudiando diseños subóptimos y sus propiedades para el problema del control multiperiodo ya indica que la obtención de diseños óptimos debe de ser prohibitivo. Estos sólo se pueden encontrar cuando el horizonte N es muy pequeño ($N = 2$ ó 3). En principio, tal como describen De Groot (1970), Berry y Fristedt (1985) o Putterman (1994) para el caso general del diseño secuencial, se podría utilizar la programación dinámica para identificar diseños óptimos, pero esto requeriría la resolución de integrales N -dimensionales muy complejas, además de minimización de funciones entre ellas. Debido a dicha complejidad no se conoce demasiado sobre estas reglas óptimas. A propósito de un problema de diseño similar para el que sólo se resuelven casos particulares mucho más sencillos que el que se trata aquí, Hardwick y Stout (1998) afirman que la implementación de la programación dinámica requiere cálculos prohibitivos incluso en el caso en el que se esté en condiciones de programar en paralelo.

En problemas semejantes sólo ha sido posible implementar diseños óptimos cuando la respuesta es Bernuolli (y, por lo tanto, el soporte del diseño son dos puntos), la distribución a priori sobre los parámetros son dos Beta's independientes y en cada estadio se elige sólo entre dos opciones tal como aparece en Berry y Eick (1995), Ginebra y Clayton (1995) y Hardwick y Stout (1998). Para el problema de control la respuesta es continua y el soporte del diseño acostumbra a ser un intervalo de R , esto es, en cualquier estadio se tiene un número infinito de

elecciones posibles. Algunas caracterizaciones parciales del diseño óptimo para control multiperiodo para el modelo lineal simple pueden encontrarse en Prescott (1972), Berliner (1982, 83, 87), Berger et al (1982), Srinivasan (1984) y en otras referencias allí citadas.

A continuación se presenta como se calcularía el diseño óptimo con horizonte $N = 2$, para pérdida cuadrática, $r_2(d)$, y modelo lineal simple

$$y_i = \mathbf{b}_0 + \mathbf{b}_1 x_i + \mathbf{e} ,$$

con la idea de mostrar hasta que punto estos cálculos son prohibitivos. Obtener el diseño óptimo para este horizonte requiere calcular (x_1^{opt}, x_2^{opt}) , es decir, hay que especificar las elecciones óptimas para cada uno de los dos estadios. El punto óptimo en el segundo estadio x_2^{opt} en función de (x_1^{opt}, y_1) y de la información a priori $\mathbf{p}_0(\mathbf{b})$ es:

$$x_2^{opt} | x_1^{opt}, y_1, \mathbf{p}_0(\mathbf{b}) = \frac{T E[\mathbf{b}_1 | x_1^{opt}, y_1, \mathbf{p}_0] - E[\mathbf{b}_0 \mathbf{b}_1 | x_1^{opt}, y_1, \mathbf{p}_0]}{E[\mathbf{b}_1^2 | x_1^{opt}, y_1, \mathbf{p}_0]} .$$

Notar que sustituyendo x_1^{opt} por un x_1 cualquiera se obtendrá la elección óptima en el segundo estadio cuando se ha observado en el primer estadio en x_1 . Para calcular el punto óptimo en el primer estadio, x_1^{opt} , se debe encontrar el punto x_1 tal que minimice:

$$E_0[(y_1(x_1) - T)^2 + (y_2(x_2^{opt}) - T)^2],$$

siendo E_0 la esperanza sobre la distribución conjunta de (y_1, \mathbf{b}) en el estadio cero. Notar que x_2^{opt} es función de x_1 , y_1 y de $\mathbf{p}_0(\mathbf{b})$. Desarrollando los cuadrados y considerando que las \mathbf{e}_i son independientes e idénticamente distribuidas con media cero, se tiene que para obtener x_1^{opt} hay que minimizar en x_1 :

$$\begin{aligned} \min_{x_1} E_0[(\mathbf{b}_0 + \mathbf{b}_1 x_1 - T)^2 + (\mathbf{b}_0 + \mathbf{b}_1 x_2^{opt} - T)^2] = \\ \min_{x_1} x_1^2 E_0[\mathbf{b}_1^2] + 2(E_0[\mathbf{b}_0 \mathbf{b}_1] - T E_0[\mathbf{b}_1]) x_1 + E_0[(\mathbf{b}_1 x_2^{opt})^2] + 2 E_0[(\mathbf{b}_0 - T) \mathbf{b}_1 x_2^{opt}], \end{aligned}$$

donde $E_0[\mathbf{b}_1^2]$, $E_0[\mathbf{b}_0\mathbf{b}_1]$ y $E_0[\mathbf{b}_1]$ son los momentos de la distribución a priori $\mathbf{p}_0(\mathbf{b})$. En general, el cálculo de $E_0[(\mathbf{b}_1x_2^{opt})^2]$, $E_0[\mathbf{b}_1x_2^{opt}]$ y $E_0[\mathbf{b}_0\mathbf{b}_1x_2^{opt}]$ pasa por la resolución de integrales muy complejas calculables sólo a través de métodos numéricos.

Como caso particular, considérese el modelo lineal normal con distribución a priori sobre los parámetros $\mathbf{p}_0(\mathbf{b})$ normal. Usando uso de los resultados distribucionales dados en la Sección 1 del Apéndice I sigue que:

$$x_2^{opt} = \frac{a_1y_1^2 + a_2y_1 + a_3}{b_1y_1^2 + b_2y_1 + b_3},$$

donde a_i y b_i para $i=1,2,3$ son coeficientes que dependen de x_1 y de los momentos de $\mathbf{p}_0(\mathbf{b})$. Así, por ejemplo, $E_0[\mathbf{b}_1x_2^{opt}]$ será la solución de la integral doble:

$$E_0[\mathbf{b}_1x_2^{opt}] = \iint_{y_1 \times \mathbf{b}} \mathbf{b}_1 \frac{a_1y_1^2 + a_2y_1 + a_3}{b_1y_1^2 + b_2y_1 + b_3} \mathbf{p}_0(\mathbf{b})h(y_1 | x_1, \mathbf{b}) d\mathbf{b} dy_1 =$$

$$\iint_{\mathbf{e}_1 \times \mathbf{b}} \mathbf{b}_1 \frac{a_1(\mathbf{b}_0 + \mathbf{b}_1x_1 + \mathbf{e}_1)^2 + a_2(\mathbf{b}_0 + \mathbf{b}_1x_1 + \mathbf{e}_1) + a_3}{b_1(\mathbf{b}_0 + \mathbf{b}_1x_1 + \mathbf{e}_1)^2 + b_2(\mathbf{b}_0 + \mathbf{b}_1x_1 + \mathbf{e}_1) + b_3} \mathbf{p}_0(\mathbf{b})g(\mathbf{e}_1) d\mathbf{b} d\mathbf{e}_1.$$

siendo $g(\mathbf{e}_1)$ la función de densidad de una $N(0, \mathbf{S}^2)$. La gran dificultad encontrada cuando el horizonte es sólo $N=2$ baticina el grado de complejidad que se presentará cuando el horizonte sea mayor.

1.6 Descripción de los contenidos.

Como se acaba de comentar la implementación de los diseños óptimos para este problema de control es prácticamente imposible incluso para modelos tan sencillos como el de regresión lineal simple. En vez de intentar alcanzar la optimalidad, en esta memoria definimos

familias de diseños construidas a base de perturbar los diseños más utilizados para este tipo de problema, que ya de por sí tienen buenas propiedades. Dentro de dichas familias se buscará aquellos diseños que minimizan el valor esperado de la función de pérdida a través de simulación de una manera análoga a la usada en Muller y Parmigianni (1993) para problemas de diseños óptimos no secuenciales y Ginebra y Clayton (1995) para un problema de diseños secuenciales distinto al de control.

Esta memoria está estructurada en tres partes. En la primera, que abarca los Capítulos 1 y 2, se presentan los contenidos generales y los diseños más utilizados para resolver el problema de control multiperiodo. En la segunda parte se proponen familias de diseños para el problema de control con una única variable de entrada, $k=1$, y abarca desde el Capítulo 3 al 5. En la tercera parte, que coincide con el Capítulo 6, se introduce una nueva familia de diseños válida para cualquier número de variables de control y que incluye, como caso particular, el de un solo control tratado en la segunda parte. A continuación se describen los contenidos de cada capítulo.

En el Capítulo 2 se estudian los diseños más utilizados para el problema de control multiperiodo. Estos son los diseños *certainty equivalence* y los diseños *miopes*, y constituirán la base para la construcción de las familias de diseños propuestas en el resto de la tesis. En este capítulo se resumirán las principales características de tales diseños y de los estimadores construidos a partir de las observaciones obtenidas de los mismos. Además se exponen versiones Bayesianas de los diseños *certainty equivalence* y *miope* para el modelo lineal múltiple, comparando el rendimiento de los mismos.

En el Capítulo 3 se proponen los diseños *Fieller*, que perturban *certainty equivalence* con el objetivo de mejorar las estimaciones de los parámetros en los estadios iniciales del diseño y con ello favorecer el control en estadios posteriores. Ellos están soportados por un intervalo que contiene la elección *certainty equivalence*. Este intervalo sirve para controlar la dirección y la magnitud de las perturbaciones a realizar sobre la elección *certainty equivalence*. Para

ejemplificar el uso de tales diseños se presentarán diversos ejemplos y se realizará un análisis de sensibilidad que refleje el comportamiento de las mejoras sobre *certainty equivalence* en función de la priori seleccionada, el horizonte considerado N y el tamaño de la región experimental C .

En el Capítulo 4 se define una nueva familia de diseños que utilizan regiones creíbles sobre la raíz para controlar el grado de desviación y la dirección de la misma respecto de la elección *certainty equivalence*. En particular, ello se ejemplifica a través del uso de regiones *HPD* y de intervalos centrales a posteriori. Al igual que con los diseños *Fieller*, las perturbaciones de los primeros estadios incentivan el aprendizaje sobre $f(x; \mathbf{b})$ y favorecen el control posterior, reduciendo de esta manera la pérdida total esperada. Este tipo de diseños no sólo es aplicable al problema de control. Es por ello que, además de dar ejemplos y realizar un análisis de sensibilidad de estos diseños para el problema de control, se ejemplifica su uso sobre el problema del *Response Surface Bandit* definido en Ginebra y Clayton (1995). Es un problema similar, pero con el propósito de modelar situaciones en las que se está interesado en que todas las respuestas y_1, \dots, y_N estén próximas al óptimo de $f(x; \mathbf{b})$.

En el Capítulo 5 se define la familia de diseños *san*, construida sobre una modificación de las secuencias de aproximación estocásticas definidas por Robbins y Monro (1951). Estas aproximaciones están ligadas a convergencias asintóticas hacia la raíz y las modificaciones realizadas están dirigidas a introducir en cada estadio la máxima cantidad de información suministrada por los datos hasta ese momento. Dicha familia contiene como miembros a los diseños *certainty equivalence* secuencial y no secuencial, y para el modelo lineal los puntos del diseño *san* pueden expresarse como promedios ponderados de las distintas elecciones *certainty equivalence* hasta ese momento. El capítulo finaliza con ejemplos de su aplicación y un análisis de sensibilidad de los mismos que refleja las mejoras sobre *certainty equivalence* en función de

la priori, horizonte y región experimental. También se apunta una posible generalización para implementar estos diseños al caso $k > 1$.

Los diseños precedentes son válidos para $k=1$ pero no son siempre fácilmente extensibles al caso $k > 1$. En el Capítulo 6, la tercera parte de esta memoria, se propone una nueva familia de diseños *híbridos* mucho más general que las anteriores propuestas, y que se puede aplicar a problemas de control con más de una variable de entrada. Su construcción se basa en la minimización de una combinación convexa de las expresiones que deben minimizarse para las elecciones *certainty equivalence* y *miope*. Este hecho hace que tanto el diseño *certainty equivalence* como el *miope* pertenezcan a tal familia, permitiendo así el poder compensar las buenas propiedades asintóticas del *certainty equivalence* con las buenas cualidades del diseño *miope* para horizontes muy reducidos. Una ventaja de esta familia respecto a las anteriores es que sólo necesita un índice, lo que facilita la obtención de la estrategia óptima dentro de dicha familia.

En el Capítulo 6 también sorprende la estrategia híbrida para el modelo lineal múltiple en cada estadio, mostrando que se trata de una combinación convexa de la elección *certainty equivalence* y de la elección *miope*. Se estudia también el caso particular del modelo lineal normal con priori conjugada donde se descubre que en cualquier estadio n , tanto si \mathbf{s}^2 se supone conocida o desconocida, los puntos de los diseños *certainty equivalence* y *miope* se hallan siempre sobre una misma recta que depende únicamente de los parámetros de la distribución a priori $\mathbf{p}_0(\mathbf{b})$ ó $\mathbf{p}_0(\mathbf{b}, \mathbf{s}^2)$ y no de las x_1, \dots, x_n escogidas ni de las y_1, \dots, y_n observadas. El capítulo termina comparando la implementación de esta familia en el problema de un único control con el caso de tres controles y analizando la sensibilidad de dichos diseños en función de los parámetros de la priori, tamaño de la región experimental y horizonte N .

Como se comentó en el epígrafe dedicado al modelo lineal, esta memoria incluye dos apéndices en los que se presentan los ingredientes necesarios para implementar todas nuestras

propuestas para el modelo lineal simple con ocho prioris distintas y para el modelo lineal múltiple con priori conjugada.

1.7 Notación.

A continuación se introducen las abreviaciones y la notación más empleada a lo largo de la redacción de esta memoria:

- A la raíz de $f(x; \mathbf{b}) = T$ cuando sólo existe un control lo denotaremos por \mathbf{q} , mientras que al máximo de $f(x; \mathbf{b})$ por \mathbf{k} .
- A la desviación típica de . lo denotaremos por $s.d.(.)$.
- A la distribución a posteriori de . en el estadio n cuando se ha elegido x_1, \dots, x_n y observado y_1, \dots, y_n partiendo de la distribución a priori \mathbf{p}_0 sobre los parámetros, $\mathbf{p}(\cdot | Y_n, X_n, \mathbf{p}_0)$, la denotaremos por $\mathbf{p}_n(\cdot)$.
- A la distribución condicional a posteriori de . dado a en el estadio n , $\mathbf{p}(\cdot | a, Y_n, X_n, \mathbf{p}_0)$, la denotaremos por $\mathbf{p}_n(\cdot | a)$.
- A los estimadores mínimos cuadrados del parámetro \mathbf{b}_i en el estadio n lo denotaremos por $\hat{\mathbf{b}}_{in}^{ls}$.
- A la moda, mediana, media y varianza a posteriori del parámetro \mathbf{b}_i en el estadio n los denotaremos por \mathbf{b}_{in}^{pmo} , \mathbf{b}_{in}^{pme} , $E_n[\mathbf{b}_i]$ y $V_n[\mathbf{b}_i]$, respectivamente.
- A la estimación de $f(x; \mathbf{b})$ que se obtiene al sustituir los parámetros por las modas a posteriori en el estadio n la denotaremos por $f_n^{pm}(x_0; \mathbf{b}) = \mathbf{b}_{in}^{pmo}(x_0 - \mathbf{q}_n^{pmo})$.
- A la normal de parámetros \mathbf{m} y \mathbf{s}^2 truncada en el intervalo $[a, b]$ la denotaremos por $N^{[a, b]}(\mathbf{m}, \mathbf{s}^2)$.

- A la función de distribución de una normal tipificada con $m=0$ y $s^2=1$ la denotaremos por $F(x)$.
- Al sumatorio de a_i para los estadios $i=1, \dots, n$ lo denotaremos por $\sum a_i$.

CAPITULO 2.

DISEÑOS *CERTAINTY EQUIVALENCE* Y DISEÑOS *MIOPE*S.

Como ya se comentó en el Capítulo 1, la obtención de diseños óptimos para el tipo de problemas expuestos es prácticamente imposible, aún incluso haciendo uso de grandes supercomputadoras. Por ello, muchos investigadores utilizan diseños basados en heurísticas *ad-hoc* que, sin ser óptimos, son razonablemente eficientes y mucho más sencillos de implementar que el óptimo.

En este capítulo se consideran las versiones Bayesianas de los conocidos diseños *certainty equivalence*, secuenciales y los no secuenciales, y de los diseños *miopes*, calculando sus expresiones para el modelo lineal múltiple. Para entender bien su funcionamiento se completa el capítulo comparando *certainty equivalence* secuencial con el no secuencial, así como *certainty equivalence* con los diseños *miopes*.

Estos diseños serán de gran utilidad en los capítulos posteriores pues se partirá de ellos a la hora de construir familias de diseños secuenciales en las que buscar un diseño que los mejore y que pueda estar más cerca del óptimo. Existe gran cantidad de artículos tratando aspectos asintóticos de tales diseños como la consistencia de los estimadores de los parámetros del modelo obtenidos a través de dichos diseños. En este trabajo nos hemos propuesto estudiar

el comportamiento de los mismos para problemas con un horizonte N finito, y a menudo pequeño. Además, en este capítulo se repasará bibliográficamente tales métodos, comentando las principales aportaciones realizadas en dichos artículos.

En las Secciones 2.1 y 2.2 se introducen los diseños *certainty equivalence* secuenciales y se comparan con una versión no secuencial. Para el modelo lineal múltiple se propone una versión Bayesiana del diseño *certainty equivalence* que es cotejada en la Sección 2.4 con la versión para este mismo modelo del diseño *miope* dado en la Sección 2.3. En la Sección 2.5 se argumenta el porqué de la construcción de los nuevos diseños que se definen en los capítulos siguientes de esta memoria sobre la base de perturbar tanto a *certainty equivalence* como a *miope*. El capítulo finaliza con un apéndice en el que se demuestran los resultados expuestos.

2.1 Diseños *certainty equivalence*.

Si $f(x; \mathbf{b})$ fuese conocida se sabrían los $x^0 = (x_1^0, \dots, x_k^0)$ que resuelven $f(x^0; \mathbf{b}) = T$, y experimentar en $x_n^d = x^0$ sería óptimo para cualquier función de pérdida que penalizara la distancia al valor objetivo T . Los diseños *certainty equivalence* secuenciales, denotados por d_{ce} , son aquéllos que serían óptimos si el estimador actual fuese $f(x; \mathbf{b})$. Dado un estimador de $f(x; \mathbf{b})$ en el estadio n , $\hat{f}_n(x; \mathbf{b})$, los diseños *certainty equivalence* observan y_{n+1} en un punto x_{n+1}^{ce} que resuelve $\hat{f}_n(x; \mathbf{b}) = T$.

En la literatura de control adaptativo, tal como en Astrom y Wittenmark (1995), es habitual que $\hat{f}_n(x; \mathbf{b}) = f(x; \hat{\mathbf{b}}_n)$, y así los diseños *certainty equivalence* son aquéllos que serían óptimos si \mathbf{b} fuera igual a un estimador de \mathbf{b} , $\hat{\mathbf{b}}_n$. Intrínseco en esta definición está el hecho de que existen tantos diseños *certainty equivalence* como posibles estimadores para la superficie. En particular, en el caso de un único control, se podría estimar la raíz \mathbf{q} a partir de la

moda, media o mediana a posteriori, y definir a través de ellos tres diseños *certainty equivalence* distintos, $x_{n+1}^{ce} = \hat{\mathbf{q}}_n$. En el contexto Bayesiano se considera que lo más coherente es estimar $f(x; \mathbf{b})$ a través de la esperanza de la distribución predictiva a posteriori de y , $\hat{f}_n(x; \mathbf{b}) = E_n[y(x)]$. En este caso x_{n+1}^{ce} será un punto que resuelve $E_n[y(x) - T] = 0$ o, equivalentemente, un punto que minimiza $(E_n[y(x) - T])^2$ en x . Otros estimadores para $f(x; \mathbf{b})$ podrían ser también utilizados.

2.1.1 Repaso bibliográfico.

Existe un gran número de referencias estadísticas sobre los diseños *certainty equivalence* para el caso de una sola variable control, cuando $k = 1$. Prescott (1972) implementa una regla *certainty equivalence* Bayesiana para el modelo lineal simple con $\mathbf{b}_0 = 0$ y estimando \mathbf{b}_{1n} por medio de mínimos cuadrados, mostrando que tal regla no es óptima. Berliner (1982, 83, 87) expone que algunas reglas *certainty equivalence* no son admisibles. Anderson y Taylor (1976) realizan un estudio experimental no Bayesiano sobre las propiedades asintóticas de la regla *certainty equivalence* definida con estimadores mínimos cuadrados en el modelo lineal simple con los dos parámetros desconocidos, probando que tal regla converge con probabilidad uno al verdadero valor de la raíz y que es asintóticamente eficiente. Sin embargo, mientras una combinación de los parámetros estimados en el problema de control es consistente, no existen evidencias confirmatorias de que las estimaciones individuales por sí solas lo sean.

Lai y Robbins (1982) también estudian el comportamiento asintótico de la regla *certainty equivalence* definida por Anderson y Taylor (1976) incluyendo la necesidad de que el signo de los estimadores de \mathbf{b}_1 fuera el mismo que el de \mathbf{b}_1 , dando una demostración matemática más precisa de la convergencia de tal regla. Ellos ya habían propuesto un contraejemplo de la pérdida de convergencia en el supuesto de que el signo de \mathbf{b}_1 fuese

diferente al de sus estimadores. Ying y Wu (1997) estudian diseños secuenciales que observan y_{n+1} en el estimador máximo verosímil de \mathbf{q} para el estadio n , $x_{n+1} = \hat{\mathbf{q}}_n^{mie}$, y muestran como éstos tiene buenas propiedades asintóticas bajo algunos modelos específicos, como modelos de localización y de localización y escala. Recientemente Chen y Hu (1998) prueban la optimalidad asintótica de los diseños *certainty equivalence* Bayes para modelos lineales simples, además, si el coste del control total tiende a infinito, lo normal generalmente, los estimadores Bayes para la pendiente y el término independiente son consistentes.

En el caso de más de una variable control y con un modelo de regresión múltiple, $y_n = \mathbf{b}_1 x_{n1} + \dots + \mathbf{b}_k x_{nk} + \mathbf{e}_n$, ha sido de gran interés establecer la consistencia fuerte de los estimadores mínimos cuadrados utilizando los puntos obtenidos con la regla *certainty equivalence*. Lai y Wei (1982) muestran que el estimador mínimos cuadrados es fuertemente consistente si el mínimo autovalor $\mathbf{I}_{min}(n)$ y el máximo autovalor $\mathbf{I}_{max}(n)$ de $Z_n' Z_n$ satisfacen

$$\mathbf{I}_{min}(n) \rightarrow \infty, \log \mathbf{I}_{max}(n) = o\{\mathbf{I}_{min}(n)\}$$

casi seguro. Rootzen y Sternby (1984) exponen que si las $\{\mathbf{e}_n\}$ son independientes normalmente distribuidas y \mathbf{b} sigue una distribución normal multivariante entonces el estimador mínimos cuadrados es fuertemente consistente si $\mathbf{I}_{min}(n) \rightarrow \infty$ casi seguro. La demostración de tal resultado requiere que la distribución a posteriori de \mathbf{b} sea normal con matriz de covarianzas esencialmente la misma que $(Z_n' Z_n)^{-1}$. Hu (1997) amplía estos resultados a errores aleatorios generales y usa las matrices de covarianzas a posteriori para estudiar la consistencia fuerte de los estimadores Bayes. En el contexto Bayesiano sigue siendo un problema abierto el hecho de que para obtener la consistencia fuerte de los estimadores Bayes sea suficiente el que se verifique o no que $\mathbf{I}_{min}(n) \rightarrow \infty$ casi seguro.

2.1.2 Implementación para modelo lineal.

A continuación se describe como se implementa el diseño *certainty equivalence* para el modelo lineal. Es importante notar la diferencia entre el caso de una sola variable de control y el de más de una.

Cuando hay más de una variable de control, es decir, cuando $k > 1$, las soluciones de $f(x; \mathbf{b}) = T$ como de $\hat{f}_n(x; \mathbf{b}) - T = 0$ definen hiperplanos y no un número real. En dicho caso y, en el contexto Bayesiano, se considerará x_{n+1}^{ce} como la solución de $E_n[y(x) - T] = 0$ con la mínima varianza de predicción, $V_n[y(x)]$. Para un único control, cuando $k = 1$, dado el estimador para $f(x; \mathbf{b})$, el diseño *certainty equivalence* secuencial queda unívocamente determinado siendo el mismo para cualquier función de pérdida. La siguiente proposición presenta la expresión de la elección *certainty equivalence* para el modelo lineal múltiple:

Proposición 2.1 Sea $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i$, $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$ y $\mathbf{b}' = (\mathbf{b}_0, \dots, \mathbf{b}_k) \sim \mathbf{p}_0(\mathbf{b})$, independiente de \mathbf{e} . El diseño *certainty equivalence* definido como la solución de $E_n[y(x) - T] = 0$ que minimiza la varianza de predicción observa y_{n+1} en

$$x_{n+1}^{ce} = V_n^{-1}[\mathbf{b}_I] \left(\frac{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] COV_n[\mathbf{b}_0, \mathbf{b}_I] + (T - E_n[\mathbf{b}_0]) E_n[\mathbf{b}_I] - COV_n[\mathbf{b}_0, \mathbf{b}_I]}{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I]} \right),$$

donde $\mathbf{b}'_I = (\mathbf{b}_1, \dots, \mathbf{b}_k)$ es el vector de todos los parámetros excepto \mathbf{b}_0 y $COV_n[\mathbf{b}_0, \mathbf{b}_I]$ es el vector $k \times 1$ de covarianzas a posteriori entre \mathbf{b}_0 y los k elementos de \mathbf{b}_I .

La demostración de la Proposición 2.1 se encuentra en el Apéndice 2.A. Notar que dicha elección depende de los momentos de la distribución a posteriori de \mathbf{b} de orden uno y dos, y que \mathbf{s}^2 no tiene que ser necesariamente conocida. El siguiente corolario muestra la expresión en el supuesto de un modelo lineal simple:

Corolario 2.1 Sea $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \mathbf{b}_1 x$, $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$, $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1) \sim \mathbf{p}_0(\mathbf{b})$, independiente de \mathbf{e} . El diseño *certainty equivalence* definido como la solución de $\hat{f}_n(x; \mathbf{b}) = E_n[y(x)] = T$ observa y_{n+1} en el $x_{n+1}^{ce} \in C$ más cercano a

$$x_{n+1} = \frac{(T - E_n[\mathbf{b}_0])}{E_n[\mathbf{b}_1]}.$$

2.1.3 Diseño *certainty equivalence* no secuencial.

Para implementar diseños totalmente secuenciales es necesario conocer la historia completa de los niveles de entrada $X'_n = (x_1, \dots, x_n)$ y observaciones $Y'_n = (y_1, \dots, y_n)$ antes de seleccionar x_{n+1} . En ocasiones esto no es factible, y a veces se deben elegir todos los x_n 's basándose en lo que se sabe de $f(x; \mathbf{b})$ al principio, de forma no secuencial. Atendiendo a la información a priori existente se definen a continuación los diseños *certainty equivalence* no secuenciales, denotados por d_{CE} .

Definición 2.1 Dado un estimador de $E[y|x; \mathbf{b}]$ basado en $\mathbf{p}_0(\mathbf{b})$, $\hat{f}_0(x; \mathbf{b})$, el diseño *certainty equivalence* no secuencial observa y_n en el punto $x^{CE} \in C$ más cercano al x_1 que resuelve $\hat{f}_0(x; \mathbf{b}) = T$.

En particular, se puede estimar $f(x; \mathbf{b})$ a través de $\hat{f}_0(x; \mathbf{b}) = E_0[f(x; \mathbf{b})]$, donde la esperanza es respecto a $\mathbf{p}_0(\mathbf{b})$, o a través de $\hat{f}_0(x; \mathbf{b}) = f(x; \hat{\mathbf{b}}_0)$ donde $\hat{\mathbf{b}}_0$ es un estimador a priori para \mathbf{b} . Notar que existe toda una gama de posibles diseños intermedios entre el secuencial y el no secuencial, según se actualice el estimador \hat{f}_m en algunos m estadios. Un aspecto importante es determinar cómo se comporta el diseño secuencial frente a su contrapartida no secuencial. En la siguiente sección se comparan estos dos diseños extremos.

En el resto de este trabajo sólo se estudiará el problema totalmente secuencial con lo que las versiones *certainty equivalence* que aparecerán a partir del Capítulo 3 van a ser secuenciales.

2.2 Comparación de *certainty equivalence* secuencial con no secuencial.

Para evaluar el comportamiento de la elección *certainty equivalence* secuencial frente a la no secuencial se define la mejora relativa del diseño secuencial frente al no secuencial como

$$rr_i(ce) = \frac{r_i(d_{CE}) - r_i(d_{ce})}{r_i(d_{ce})},$$

con d_{CE} el *certainty equivalence* no secuencial y d_{ce} el secuencial y para las pérdidas esperadas $r_i(d)$. Recordar que $r_2(d) = E[\sum_{i=1}^N (y_i(x_i) - T)^2]$, que $r_1(d)$ es igual pero sumando desviaciones en valor absoluto y $r_3(d)$ es una pérdida cuadrática en dos etapas.

La Figura 2.1 presenta los gráficos de contorno para $rr_i(ce)$ tanto para la versión conjugada como para una no conjugada del modelo lineal simple descrita en el Apéndice I y Apéndice II, respectivamente. Las pérdidas corresponden a las tres descritas en el primer capítulo, con $N = 25$, $S = 200$, $T = 0$ y $C = [-1, 1]$. Los gráficos de contorno en la primera columna son para la priori conjugada $\mathbf{b}_0 \sim N(\mathbf{m}_{b_0} = 0., \mathbf{s}_{b_0}^2)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1} = 1., \mathbf{s}_{b_1}^2 = 2.^2)$ y \mathbf{b}_1 independiente de \mathbf{b}_0 , con x_{n+1}^{ce} definido como el punto en C más cercano a $x_{n+1} = -E_n[\mathbf{b}_0]/E_n[\mathbf{b}_1]$. Los gráficos de la segunda columna son para la priori no conjugada $\mathbf{q} \sim U(-rtu, rtu)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1} = 1., \mathbf{s}_{b_1}^2 = .2^2)$ y \mathbf{b}_1 independiente de \mathbf{q} , donde x_{n+1}^{ce} es el punto en el soporte de \mathbf{q} más cercano a $x_{n+1} = \mathbf{q}_n^{pmo}$.

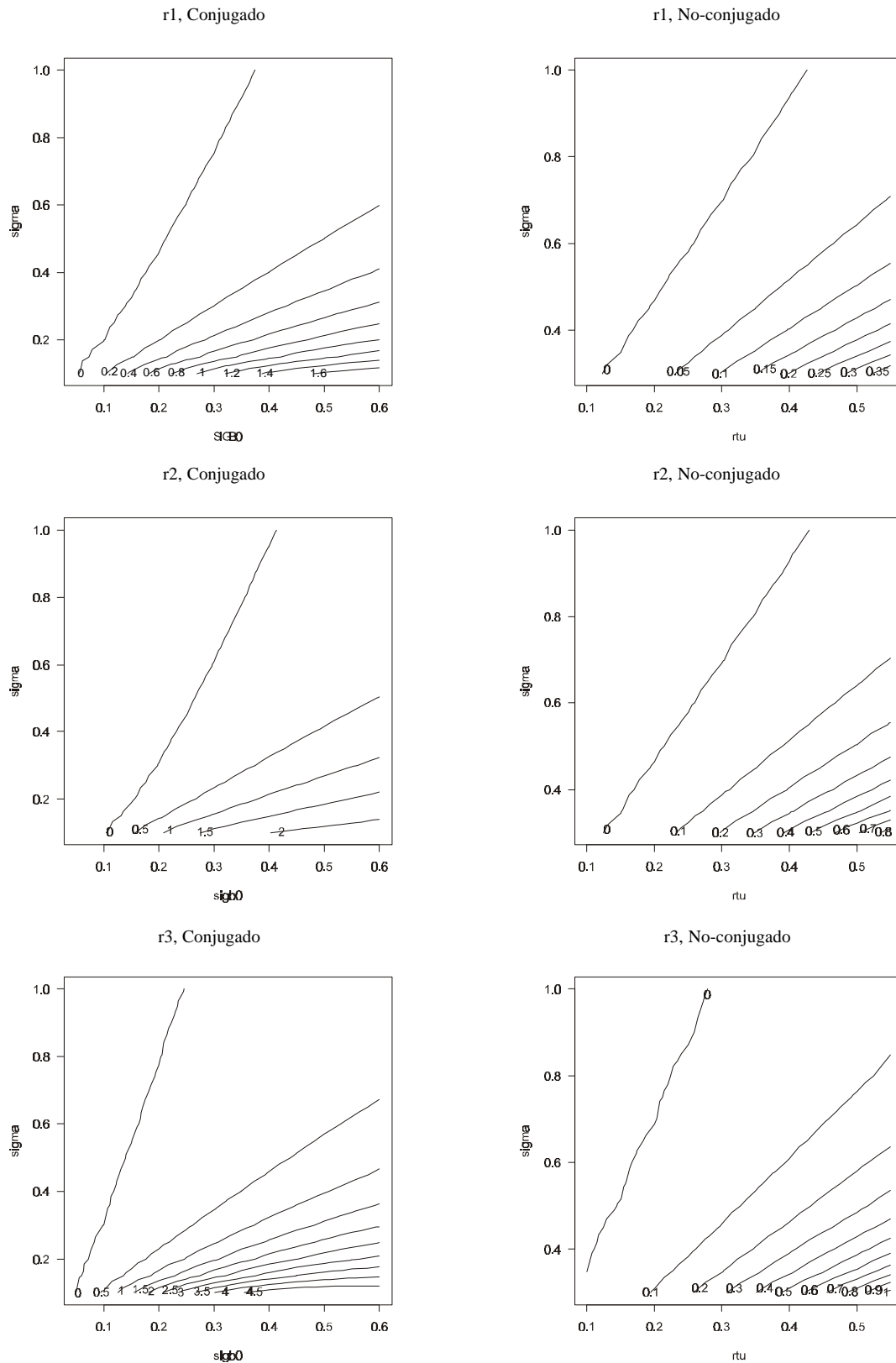


Figura 2.1. Gráfico de contorno de la mejora relativa de *certainty equivalence* secuencial frente a no secuencial, $rr_i(ce)$, para el riesgo $r_i(d)$, $i = 1, 2, 3$, con $N = 25$, $S = 200$, $T = 0$ y $C = [-1, 1]$. La primera columna es para $\mathbf{b}_0 \sim N(\mathbf{m}_{\mathbf{b}_0} = 0., \mathbf{s}_{\mathbf{b}_0}^2)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{\mathbf{b}_1} = 1., \mathbf{s}_{\mathbf{b}_1}^2 = 2.^2)$, \mathbf{b}_1 independiente de \mathbf{b}_0 . En la segunda columna $\mathbf{q} \sim U(-rtu, rtu)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{\mathbf{b}_1} = 1., \mathbf{s}_{\mathbf{b}_1}^2 = .2^2)$ y \mathbf{b}_1 independiente de \mathbf{q} .

La Figura 2.1 muestra que existen algunos casos donde mantener $x_n = x_0$ fijo como lo hace la elección no secuencial es mejor que ser secuencial. Para un $\mathbf{s}_{b_0}^2$ dado, o *rtu*, cuanto mayor es \mathbf{s}^2 más difícil es aprender secuencialmente sobre \mathbf{q} , y cuanto más al borde inferior derecho de los gráficos mucho mejor es ser secuencial. Esto se debe al efecto de tener una región C acotada, que restringe la posibilidad de que los puntos se alejen en demasía sobre todo cuando \mathbf{s}^2 es relativamente grande. Por otro lado, dado \mathbf{s}^2 , para los mayores valores de $\mathbf{s}_{b_0}^2$, o *rtu*, es decir, cuanto menos se conoce al principio de \mathbf{q} , mayor ganancia se obtiene usando lo que se aprende secuencialmente. Se ha encontrado que estas diferencias son bastante insensibles a los cambios en $\mathbf{s}_{b_1}^2$ y \mathbf{m}_{b_1} .

2.3 Diseños *miopes*.

Otro tipo de diseños secuenciales muy utilizados son los diseños *miopes*, que son los que serían óptimos si quedara sólo una observación por realizar y, por tanto, sería el método a utilizar cuando el aprendizaje no fuera necesario. Como tal, son eficientes para problemas con horizontes N pequeños y moderados. A diferencia de los *certainty equivalence*, los diseños *miopes* no requieren de la especificación de un estimador para $f(x; \mathbf{b})$, además quedan definidos unívocamente para cualquier k y sí dependen de la función de pérdida especificada en el problema. En el caso particular en el que se minimiza $r_2(d) = E[\sum_{i=1}^N (y_i(x_i) - T)^2]$, el diseño *miope* observa y_{n+1} en el punto x_{n+1}^{my} que minimiza $E_n[(y(x) - T)^2]$ en x .

Notar que cuando para $k > 1$ se define x_{n+1}^{ce} como la solución de $E_n[y(x) - T] = 0$ con la mínima varianza de predicción, se está considerando la elección *miope* bajo pérdida cuadrática de entre todas las soluciones a la ecuación *certainty equivalence* que satisfacen

$E_n[y(x) - T] = 0$. En el caso en que se minimice $r_1(d) = E[\sum_{i=1}^N |y_i(x_i) - T|]$, la elección *miope* coincidirá con la elección *certainty equivalence* definida como la solución de $E_n[y(x) - T] = 0$, pero para otras funciones de pérdida en general no se verificará.

Berliner (1987) y Verdinelli y Wynn (1988) muestran cómo computar la selección *miope* no Bayesiana para casos particulares de modelos lineales múltiples. Prescott (1974) realiza un estudio comparativo en un contexto Bayesiano entre las estrategias *certainty equivalence*, *miope* (denominada por él regla de decisión adaptativa) y óptima, para un modelo lineal simple con $\mathbf{b}_0 = 0$ y horizonte $N = 2$, donde la estrategia óptima se puede obtener de forma sencilla por *backward induction* y métodos numéricos.

Lai y Robbins (1982) prueban que la elección *miope* Bayes para un modelo lineal simple donde la pendiente \mathbf{b}_1 fuese una constante conocida es optimal. En ese caso, la elección *certainty equivalence* tomando valor esperado también coincidirá con la *miope*. La siguiente proposición muestra la elección *miope* en el supuesto de un modelo lineal múltiple de la forma $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i$, utilizando el riesgo cuadrático $r_2(d)$.

Proposición 2.2 Sea $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i$, $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$ y $\mathbf{b}' = (\mathbf{b}_0, \dots, \mathbf{b}_k) \sim \mathbf{p}_0(\mathbf{b})$, independiente de \mathbf{e} . El diseño *miope* considerando riesgo cuadrático observa y_{n+1} en

$$x_{n+1}^{my} = V_n^{-1}[\mathbf{b}_I] \left(\frac{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] COV_n[\mathbf{b}_0, \mathbf{b}_I] + (T - E_n[\mathbf{b}_0]) E_n[\mathbf{b}_I] - COV_n[\mathbf{b}_0, \mathbf{b}_I]}{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I] + 1} \right),$$

donde $\mathbf{b}'_I = (\mathbf{b}_1, \dots, \mathbf{b}_k)$ es el vector de todos los parámetros excepto \mathbf{b}_0 y $COV_n[\mathbf{b}_0, \mathbf{b}_I]$ es el vector $k \times 1$ de covarianzas a posteriori entre \mathbf{b}_0 y los k elementos de \mathbf{b}_I .

Tanto la Proposición 2.1 como la 2.2 se obtendrán como un caso particular de la Proposición 6.1 al sustituir el parámetro H de la misma por 0 y 1 respectivamente. Esto convertirá su demostración en equivalente a la dada en el Apéndice 6.A.2 para la Proposición

6.1 y no se presentará en este capítulo. A continuación se detalla la expresión de la elección *miope* para el caso particular de un modelo lineal simple bajo riesgo cuadrático.

Corolario 2.2 Sea $E[y | x, \mathbf{b}] = \mathbf{b}_0 + \mathbf{b}_1 x$, $V[\mathbf{e} | x, \mathbf{b}] = \mathbf{s}^2$, $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1) \sim \mathbf{p}_0(\mathbf{b})$, independiente de \mathbf{e} . El diseño *miope* considerando riesgo cuadrático observa y_{n+1} en el $x_{n+1}^{my} \in C$ más cercano a

$$x_{n+1} = \frac{T E_n[\mathbf{b}_1] - E_n[\mathbf{b}_0 \mathbf{b}_1]}{E_n[\mathbf{b}_1^2]}.$$

2.4 Comparación de *certainty equivalence* frente a *miope*.

Por propia definición de las estrategias *certainty equivalence* y *miope*, y por las referencias bibliográficas ya comentadas para ambos diseños, está claro que si el horizonte N es pequeño la estrategia *miope* estará muy cerca de la óptima, y según aumenta N se va alejando de la óptima acercándose hacia ésta la *certainty equivalence*. Como se detalló en la Sección 2.3, la expresión de la elección *certainty equivalence* para el modelo lineal múltiple dada en la Proposición 2.1 coincide con la *miope* que obtuvimos en la Proposición 2.2 al sustituir el uno del denominador por cero. Como explicaremos en el Capítulo 6, este hecho será aprovechado para crear una familia de diseños *híbridos* indexada por un parámetro H que contiene a ambos. Por tanto, las gráficas presentadas en ese capítulo servirán para la comparación de ambas estrategias para diferentes priori y distintas incertidumbres sobre el modelo, dependiendo de si el parámetro H está más próximo a cero o a uno.

En las Figuras 6.3 y 6.4 (Capítulo 6) se observará que los diseños *miope* se comportan bastante bien para horizontes pequeños, empeorando al aumentar N , con la peculiaridad de verse

poco afectados por el tamaño de la región experimental C . En cambio, el diseño *certainty equivalence* mejora según aumenta N , a la vez que es muy sensible a las modificaciones en el tamaño de C , empeorando su eficacia según aumenta éste. También se sigue que para una N y un tamaño de región experimental C dado, *certainty equivalence* mejorará cuanto menos se sepa sobre \mathbf{b}_1 , y menor sea \mathbf{s}^2 , aunque en general comprobaremos que el *miope* tendrá mejores prestaciones que el *certainty equivalence*.

2.5 Necesidad de perturbar *miope* y *certainty equivalence*.

Ya se ha comentado que el diseño *certainty equivalence* Bayes converge asintóticamente al óptimo y que los estimadores Bayes de los parámetros en el modelo de regresión son consistentes. Cuando el horizonte N es de tamaño moderado o pequeño, será necesario perturbar tales diseños con la idea de mejorar la eficiencia del diseño. Por otro lado, cuando N es pequeño el diseño *miope* será prácticamente óptimo. Todas estas consideraciones hacen pensar que es posible construir diseños con buenas propiedades de convergencia a partir de la base de ambos diseños.

Recordemos que para resolver el problema de control hay que encontrar el justo equilibrio entre minimizar la pérdida inmediata y maximizar el aprendizaje sobre los parámetros que permita mejorar el control futuro. La opción *miope* es la que minimiza la pérdida inmediata. Perturbando dicha opción nos permitirá inducir aprendizaje adicional. Por simulación calcularemos el grado de perturbación necesario para nuestro problema.

Así, se podría intentar construir diseños observando en puntos que se alejan de la elección *certainty equivalence*, sobre todo en los estadios iniciales del experimento, para mejorar los estimadores intermedios de los parámetros del modelo, pero de manera que en los

estadios finales se acerque a la elección *certainty equivalence*, que dará lugar a un mejor control. El problema es decidir cuánto y hacia dónde se debe observar con respecto a la elección *certainty equivalence*. En esta memoria se presentan distintas alternativas, todas ellas con la idea de construir familias de diseños que contengan a los propios diseños *certainty equivalence*, y que sepan introducir las perturbaciones ya comentadas anteriormente.

En el Capítulo 3 se definen los diseños *Fieller*, que permiten observar en puntos dentro de un intervalo creíble aproximado para la raíz q . En el Capítulo 4, se introducen familias de diseños basadas en regiones creíbles al estilo de la propuesta del Capítulo 3, pero utilizando regiones *HPD* o intervalos centrales a posteriori sobre la raíz q . En el Capítulo 5, los diseños *san* modifican las secuencias de aproximación estocástica introducidas por Robbins y Monro (1951) para generar una familia de diseños que contiene tanto al diseño *certainty equivalence* secuencial como al no secuencial, barriendo así problemas con alta incertidumbre a priori, con necesidad de ser secuencial, o con problemas donde la incertidumbre es poca y casi con la información a priori es suficiente. Aunque todas estas propuestas podrían intentar adaptarse a problemas con más de una variable de control, no se ha realizado. En cambio, en el Capítulo 6 se construye una familia de diseños que contiene tanto a *certainty equivalence* como a *miope*, y que podrá ser utilizado para problemas con más de una variable de control. El hecho de contener a estos dos diseños permitirá abordar con efectividad problemas tanto con horizontes pequeños así como con horizontes más grandes.

Apéndice 2.A Demostración Proposición 2.1.

Antes de la demostración de la Proposición 2.1 presentamos un lema dado en Tewarson (1973) que es de gran utilidad a la hora de expresar las submatrices de una matriz en función de las submatrices correspondientes de su matriz inversa, y se utilizará también para la demostración del Lema 6.2 en el capítulo dedicado a los diseños *híbridos*.

Lema 2.1 (Tewarson, 1973). Sean A y A^{-1} una matriz cuadrada $N \times N$ y su inversa, respectivamente. Supóngase A particionada como

$$A = \begin{pmatrix} P & Q \\ R & S \end{pmatrix},$$

donde P y S son matrices cuadradas de $p \times p$ y $s \times s$ respectivamente ($p + s = N$). Si la matriz A^{-1} es particionada de la misma manera,

$$A^{-1} = \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix},$$

entonces, \tilde{P} , \tilde{Q} , \tilde{R} y \tilde{S} pueden ser expresadas como

$$\begin{aligned} \tilde{P} &= (P - QS^{-1}R)^{-1}, \\ \tilde{Q} &= -(P - QS^{-1}R)^{-1}(QS^{-1}), \\ \tilde{R} &= -(S^{-1}R)(P - QS^{-1}R)^{-1}, \\ \tilde{S} &= S^{-1} + (S^{-1}R)(P - QS^{-1}R)^{-1}(QS^{-1}), \end{aligned}$$

o de forma equivalente

$$\begin{aligned} \tilde{P} &= P^{-1} + (P^{-1}Q)(S - RP^{-1}Q)^{-1}(RP^{-1}), \\ \tilde{Q} &= -(P^{-1}Q)(S - RP^{-1}Q)^{-1}, \\ \tilde{R} &= -(S - RP^{-1}Q)^{-1}(RP^{-1}), \\ \tilde{S} &= (S - RP^{-1}Q)^{-1}. \end{aligned}$$

Demostración Proposición 2.1.

Recuérdese que $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i$, $V[\mathbf{e} | x, \mathbf{b}] = \mathbf{s}^2$ y $\mathbf{b}' = (\mathbf{b}_0, \dots, \mathbf{b}_k) \sim \mathbf{p}_0(\mathbf{b})$ independiente de \mathbf{e} . Para calcular la solución de $E_n[y(x) - T] = 0$ que minimiza la varianza de predicción se debe resolver el sistema

$$\left. \begin{array}{l} \min V_n[y(x) - T] \\ \text{s.a. } E_n[y(x) - T] = 0 \end{array} \right\} \Rightarrow \left. \begin{array}{l} \min z' V_n[\mathbf{b}] z \\ \text{s.a. } z' E_n[\mathbf{b}] - T = 0 \end{array} \right\} \Leftrightarrow \min z' V_n[\mathbf{b}] z + \mathbf{I} (z' E_n[\mathbf{b}] - T),$$

con \mathbf{I} el multiplicador de Lagrange y $z' = (1, x_1, \dots, x_k)$. Derivando con respecto a \mathbf{I} y con respecto a x_i e igualando a cero se tiene que

$$\begin{pmatrix} 0 & E_n[\mathbf{b}_I]' \\ E_n[\mathbf{b}_I] & 2V_n[\mathbf{b}_I] \end{pmatrix} \begin{pmatrix} \mathbf{I} \\ x_1 \\ \vdots \\ x_k \end{pmatrix} = \begin{pmatrix} -(E_n[\mathbf{b}_0] - T) \\ -2COV_n[\mathbf{b}_0, \mathbf{b}_I] \end{pmatrix}.$$

Usando los resultados de inversión de matrices particionadas dados en el Lema 2.1 se sigue que

$$\begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = (\tilde{R} : \tilde{S}) \begin{pmatrix} -(E_n[\mathbf{b}_0] - T) \\ -2COV_n[\mathbf{b}_0, \mathbf{b}_I] \end{pmatrix}$$

con

$$\tilde{R} = \frac{V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I]}{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I]} \quad \text{y} \quad \tilde{S} = \frac{1}{2} \left(V_n^{-1}[\mathbf{b}_I] - \frac{V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I] E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I]}{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I]} \right),$$

lo que prueba la proposición. ■

Parte II:

Una sola variable de control

CAPITULO 3.

DISEÑOS *FIELLER*.

Los diseños *certainty equivalence* no consideran la incertidumbre en la estimación de la superficie. En este capítulo se perturba dicha elección construyendo diseños que dependen tanto de $\hat{y}_n(x) = \hat{f}_n(x; \mathbf{b})$ como de la desviación estándar de $\hat{y}_n(x)$, denotada por $s.d.(\hat{y}_n(x))$. Cuando al experimentar uno se aparta de x_{n+1}^{ce} se puede incurrir en una pérdida inmediata extra que puede contribuir a aprender más sobre \mathbf{q} que vía el aprendizaje accidental de *certainty equivalence* y así incrementar la posibilidad de experimentar más cerca de \mathbf{q} en los estadios posteriores. La cuestión es a qué distancia de x_{n+1}^{ce} se debe experimentar y en qué dirección nos debemos mover con respecto a la elección *certainty equivalence* con el fin de inducir la cantidad “correcta” de aprendizaje “activo” sin incurrir en una pérdida inmediata demasiado grande.

Para responder a las preguntas de cuán lejos y en qué dirección debemos movernos, en la Sección 3.2 se define una familia de diseños para el problema de control con una sola variable de entrada, indexados a través de dos índices, K y \mathbf{g} . Dichos diseños, a los que llamaremos diseños *Fieller*, se generarán de forma constructiva sobre dichos parámetros y el modelo considerado. Estos están basados en unas aproximaciones Bayesianas de los intervalos *Fieller* que se definen en la Sección 3.1. En la Sección 3.3 se ilustra el uso de dichos diseños sobre el modelo lineal,

presentando gráficos de la pérdida esperada en función de los valores de (K, \mathbf{g}) para distintas prioris.

En una segunda etapa se determina el mejor de los diseños *Fieller* dentro de la familia. Para ello se simula el uso de los diseños para un K y un \mathbf{g} determinado sobre valores de la respuesta $(y|x)$, generados a partir de la distribución predictiva a priori, y se buscan los índices (K, \mathbf{g}) que minimizan la pérdida esperada tal como se muestra en la Sección 3.4. En la Sección 3.5 se incluyen comentarios sobre posibles mejoras y ampliaciones de tales diseños, terminando el capítulo con un apéndice donde se recogen resultados particulares sobre el cálculo de los intervalos *Fieller* para el modelo lineal.

3.1 Los intervalos *Fieller*.

Uno de los métodos más utilizados para construir regiones de confianza para \mathbf{q} y muy empleados en problemas de calibración, Buonaccorsi (1986), y de regresión inversa fue propuesto por Fieller (1940) a partir de unos intervalos que llevan su nombre. Dichos intervalos están descritos con detalle en Graybill (1976) y Brown (1966). A continuación se introduce una versión Bayesiana de los mismos.

Dado $K \geq 0$, se definen $U_{n+1}(K)$ y $V_{n+1}(K)$ como las raíces de

$$\hat{y}_n(x) - K \hat{s.d.}(\hat{y}_n(x)) = T$$

e

$$\hat{y}_n(x) + K \hat{s.d.}(\hat{y}_n(x)) = T,$$

respectivamente. Esto es, $U_{n+1}(K)$ y $V_{n+1}(K)$ son las raíces de las curvas que definen los extremos inferior y superior de un intervalo de confianza aproximado para la superficie prevista en el estadio n . En el caso de que una de las dos ecuaciones tuviese más de una raíz, se elige la

más cercana al x_{n+1} solución de $\hat{y}_n(x) = T$. Si una de las ecuaciones, o ambas, no tiene raíz e \hat{y}_n es creciente, se definen los $U_{n+1}(K)$ ó $V_{n+1}(K)$ correspondientes como $+\infty$ ó $-\infty$, respectivamente, y cuando \hat{y}_n es decreciente como $-\infty$ ó $+\infty$.

Cuando tanto $U_{n+1}(K)$ como $V_{n+1}(K)$ existen, el intervalo $(V_{n+1}(K), U_{n+1}(K))$ funciona como un intervalo a posteriori para \mathbf{q} , análogo a los intervalos fiduciales descritos con detalle en Fieller (1954), y su anchura es usada en un contexto diferente por Abdelbasit y Plackett (1983) y Sitter y Wu (1993) para definir un criterio para diseños no secuenciales cuando el objetivo es estimar \mathbf{q} . La existencia de un intervalo $(V_{n+1}(K), U_{n+1}(K))$ con extremos finitos para una K determinada se puede relacionar con el rechazo de la hipótesis nula de que $\mathbf{b}_1 = 0$.

Además se definen $U'_{n+1}(K)$ y $V'_{n+1}(K)$ como los puntos en la región experimental C más cercanos a $U_{n+1}(K)$ y $V_{n+1}(K)$, respectivamente. Notar que el ancho de $[V'_{n+1}(K), U'_{n+1}(K)]$ es monótonamente creciente con K y que cuando $K = 0$ dicho intervalo degenera en el punto x_{n+1}^{ce} . Los diseños *Fieller* inducirán una acción de sondeo entorno a la elección *certainty equivalence* a base de experimentar en puntos del intervalo $[V'_{n+1}(K), U'_{n+1}(K)]$.

Las Figuras 3.1 y 3.2 ilustran todas estas definiciones para el modelo lineal simple con $C = [-1, 1]$. En la Figura 3.1 x_{n+1}^{ce} coincide con la raíz de $\hat{y}_n(x)$, es decir, dicha raíz cae dentro de la región C . En la Figura 3.1.a tanto $U_{n+1}(K)$ como $V_{n+1}(K)$ son finitos y pertenecen a C , coincidiendo con $U'_{n+1}(K)$ y $V'_{n+1}(K)$, respectivamente, mientras que en la Figura 3.1.b $U_{n+1}(K)$ está fuera de la región C y, por ello, $U'_{n+1}(K)$ coincide con el extremo de dicha región, $U'_{n+1}(K) = 1$. En los gráficos de la Figura 3.2 se ilustran dos situaciones donde x_{n+1}^{ce} coincide con un extremo de la región C debido a que la solución de $\hat{y}_n(x) = 0$ no pertenece a C . en el Gráfico 3.2.a $U_{n+1}(K)$ es finito pero está fuera de la región C y, por lo tanto, $U'_{n+1}(K) = 1$, y en 3.2.b la ecuación $\hat{y}_n(x) - K \hat{s.d.}(\hat{y}_n(x)) = 0$ no tiene solución con lo que $U_{n+1}(K) = +\infty$ y $U'_{n+1}(K) = 1$.

3.2 Los diseños *Fieller*.

Tal como se ha descrito en la introducción de este capítulo, nuestra propuesta tiene dos etapas muy diferenciadas. En una primera se define la familia de diseños *Fieller* y en una segunda se busca el mejor diseño de la familia para un problema dado.

3.2.1 Definición de la familia *Fieller*.

Los diseños *Fieller* secuenciales estarán indexados por los parámetros $(K, \mathbf{g}) \in [0, \infty) \times [-1, 1]$ y serán denotados por $d_F(K, \mathbf{g})$. Ellos observan y_{n+1} en un punto $x_{n+1}^F(K, \mathbf{g})$ en el intervalo $(V'_{n+1}(K), U'_{n+1}(K))$. El índice K regula el ancho del intervalo. Cuanto mayor sea K , más lejos de la elección *certainty equivalence* podrá observar el diseño *Fieller*. El índice \mathbf{g} indica la posición relativa de $x_{n+1}^F(K, \mathbf{g})$ en el intervalo y su signo dicta en qué dirección nos movemos con respecto a x_{n+1}^{ce} .

Dado que x_{n+1}^{ce} siempre pertenece a $[V'_{n+1}(K), U'_{n+1}(K)]$, éste divide el intervalo en dos subintervalos. Cuando $\mathbf{g} > 0$, $d_F(K, \mathbf{g})$ experimenta sobre el subintervalo más ancho, a una distancia de x_{n+1}^{ce} igual a \mathbf{g} veces el largo de ese subintervalo. Cuando $\mathbf{g} < 0$, $d_F(K, \mathbf{g})$ experimenta sobre el subintervalo más estrecho a una distancia de x_{n+1}^{ce} igual a \mathbf{g} veces el largo de ese subintervalo, como se aprecia en las Figuras 3.1 y 3.2. Así para un K fijo, cuanto más grande es $|\mathbf{g}|$ más lejos de x_{n+1}^{ce} estará $x_{n+1}^F(K, \mathbf{g})$. Cuando $|\mathbf{g}| = 1$, $d_F(K, \mathbf{g})$ experimenta en

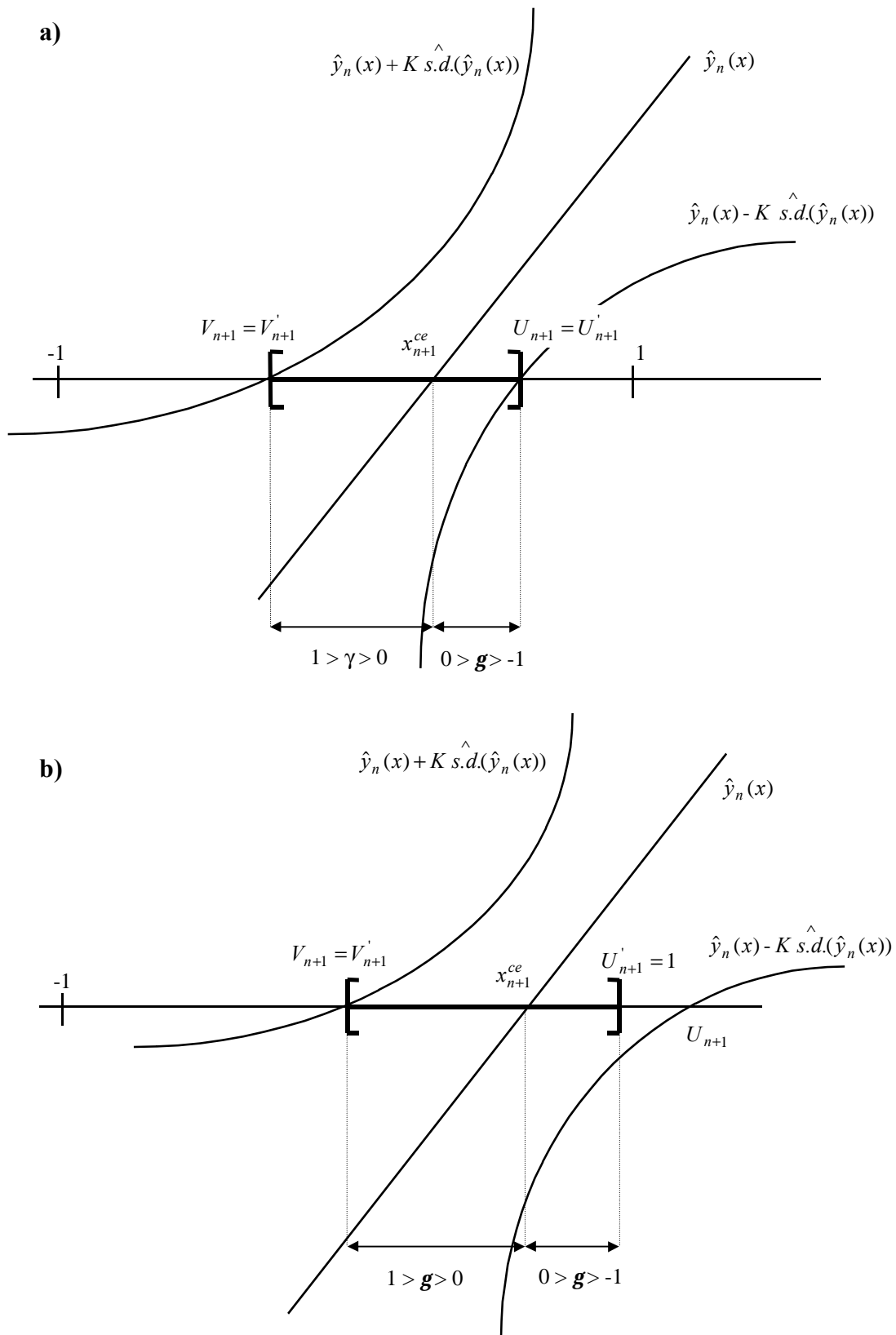


Figura 3.1. Ejemplos de intervalos $(V'_{n+1}(K), U'_{n+1}(K))$ y x_{n+1}^{ce} con $C = [-1, 1]$. En el gráfico a) $V_{n+1}(K)$ y $U_{n+1}(K)$ pertenecen a C , y en b), $U_{n+1}(K)$ queda fuera de C siendo $U'_{n+1}(K) = 1$.

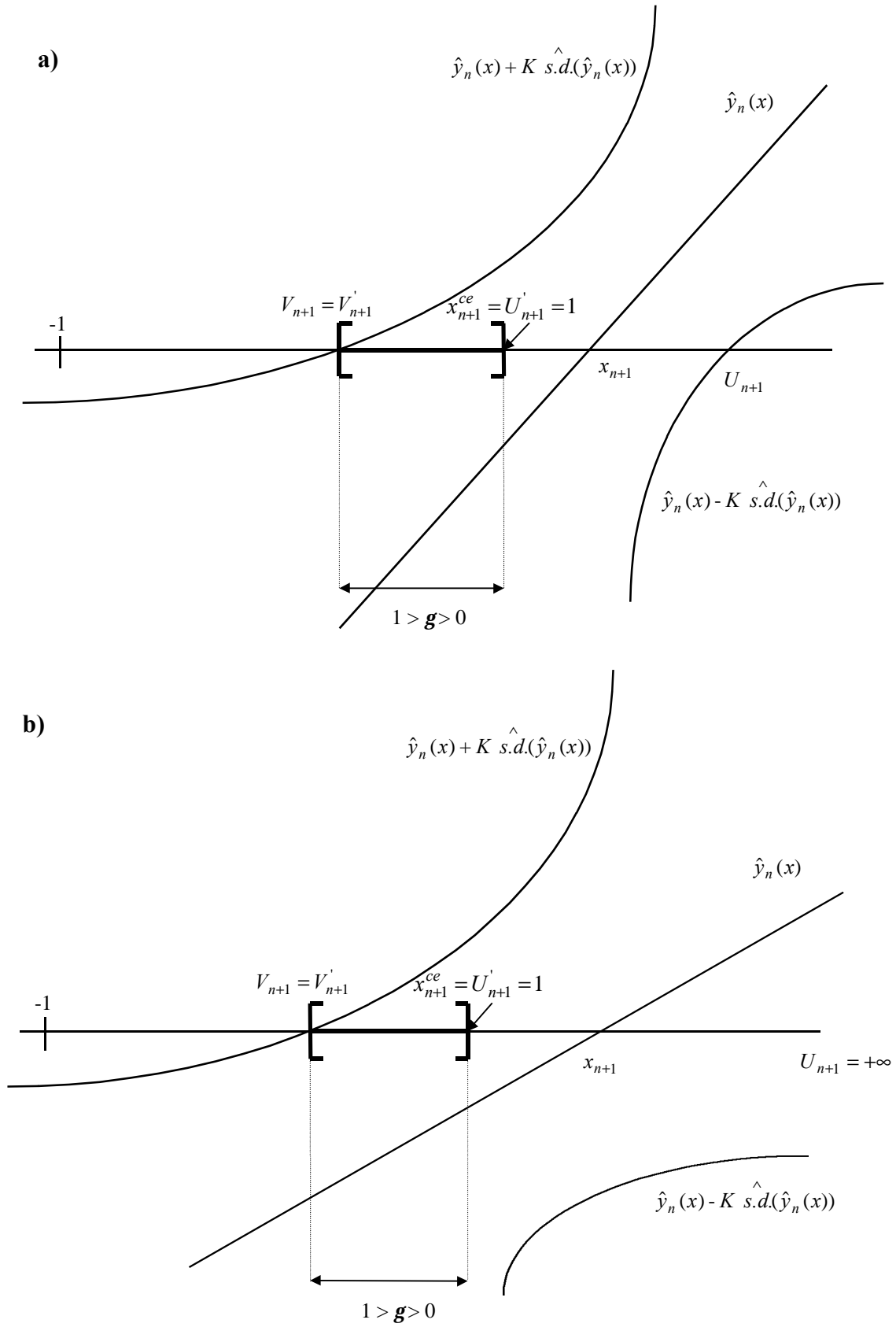


Figura 3.2. Ejemplos de intervalos (V_{n+1}', U_{n+1}') y x_{n+1}^{ce} cuando éste cae en un extremo de la región $C = [-1, 1]$. En el gráfico a) $V_{n+1}(K)$ y $U_{n+1}(K)$ son finitos y en b), $U_{n+1}(K)$ queda definido como $+\infty$ y por tanto $U_{n+1}'(K) = 1$.

$U'_{n+1}(K)$ ó $V'_{n+1}(K)$ dependiendo del signo de \mathbf{g} . La definición rigurosa de los diseños *Fieller* descrita en este párrafo es la siguiente.

Definición 3.1 Dado un estimador de $E[y|x; \mathbf{b}]$, $\hat{y}_n(x) = \hat{f}_n(x; \mathbf{b})$ y $(K, \mathbf{g}) \in [0, \infty) \times [0, 1]$, el diseño $d_F(K, \mathbf{g})$ observa y_{n+1} en

$$x_{n+1}^F(K, \mathbf{g}) = x_{n+1}^{ce} + \mathbf{g} (U'_{n+1}(K) - x_{n+1}^{ce}) I_{[|U'_{n+1}(K) - x_{n+1}^{ce}| > |V'_{n+1}(K) - x_{n+1}^{ce}|]} + (V'_{n+1}(K) - x_{n+1}^{ce}) I_{[|U'_{n+1}(K) - x_{n+1}^{ce}| < |V'_{n+1}(K) - x_{n+1}^{ce}|]},$$

y para $(K, \mathbf{g}) \in [0, \infty) \times [-1, 0]$, $d_F(K, \mathbf{g})$ observa y_{n+1} en

$$x_{n+1}^F(K, \mathbf{g}) = x_{n+1}^{ce} + \mathbf{g} (x_{n+1}^{ce} - V'_{n+1}(K)) I_{[|U'_{n+1}(K) - x_{n+1}^{ce}| > |V'_{n+1}(K) - x_{n+1}^{ce}|]} + \mathbf{g} (x_{n+1}^{ce} - U'_{n+1}(K)) I_{[|U'_{n+1}(K) - x_{n+1}^{ce}| < |V'_{n+1}(K) - x_{n+1}^{ce}|]},$$

para $n=0, 1, \dots, N-2$ y con $x_{n+1}^{ce} \in C$ igual al punto más cercano al x_{n+1} que resuelve $\hat{y}_n(x) = T$. $d_F(K, \mathbf{g})$ observa y_N en $x_N^F(K, \mathbf{g}) = x_N^{ce}$.

De acuerdo con dicha definición, cuando $(V'_{n+1}(K), U'_{n+1}(K))$ es simétrico respecto a x_{n+1}^{ce} , $|U'_{n+1}(K) - x_{n+1}^{ce}| = |V'_{n+1}(K) - x_{n+1}^{ce}|$, y así $x_{n+1}^F(K, \mathbf{g}) = x_{n+1}^{ce}$. Esto ocurre a menudo en el primer estadio debido a que la información a priori, el modelo y la región C , son tales que no priorizan ningún punto. Notar también que cuando el x_{n+1} que resuelve $\hat{y}_n(x) = T$ cae fuera de C , x_{n+1}^{ce} estará en la frontera de C y será igual a $U'_{n+1}(K)$ ó a $V'_{n+1}(K)$. En ese caso, de acuerdo con la Definición 3.1, cuando $\mathbf{g} > 0$, $d_F(K, \mathbf{g})$ experimenta en el interior de C , y cuando $\mathbf{g} < 0$ experimenta en $x_{n+1}^F(K, \mathbf{g}) = x_{n+1}^{ce}$, como se puede apreciar en la Figura 3.2.

Es importante destacar que $x_{n+1}^{ce} = x_{n+1}^F(K=0, \mathbf{g}) = x_{n+1}^F(K, \mathbf{g}=0)$ y así cuando $K=0$ ó $\mathbf{g}=0$, $d_F(K, \mathbf{g})$ coincide con el diseño *certainty equivalence*. Es reseñable observar también que cuando n aumenta, el intervalo $[V'_{n+1}(K), U'_{n+1}(K)]$ tenderá a estrecharse, puesto que $\hat{s}.d.(\hat{y}_n(x))$ tiende a disminuir y, por lo tanto, $x_{n+1}^F(K, \mathbf{g})$ se acercará más a x_{n+1}^{ce} , coincidiendo con la

necesidad decreciente de perturbar *certainty equivalence* para aprender sobre \mathbf{q} a medida que n se aproxima a N .

3.2.2 Búsqueda del mejor diseño *Fieller*.

Para encontrar el diseño $d_F(K^{i*}, \mathbf{g}^{i*})$ que minimiza $r_i(K, \mathbf{g}) = r_i(d_F(K, \mathbf{g}))$, se procede a través de simulación de MonteCarlo a partir de la distribución conjunta de los parámetros y de la respuesta y tal como sigue. Para cada (K, \mathbf{g}) se simulan repetidamente valores de los parámetros \mathbf{b} a partir de $\mathbf{p}_0(\mathbf{b})$ y se simula el uso del diseño $d_F(K, \mathbf{g})$ sobre datos procedentes del modelo para y conocido \mathbf{b} , estimando su pérdida esperada a través del promedio de las pérdidas observadas.

Entre los distintos $d_F(K, \mathbf{g})$ se encuentra el $d_F(K^{i*}, \mathbf{g}^{i*})$ que se estima sea el “mejor” para el problema, adaptando esta familia de diseños a la pérdida y al modelo considerado. Se sabe que $d_F(K^{i*}, \mathbf{g}^{i*})$ siempre mejora a los diseños *certainty equivalence*. De hecho, cuando K^{i*} ó \mathbf{g}^{i*} están cerca de cero el diseño *certainty equivalence* será el mejor de todos los diseños *Fieller*.

En la Sección 3.3 se estudia el comportamiento de las pérdidas esperadas, $r_i(K, \mathbf{g})$, $i=1,2,3$, para el modelo de regresión simple en función de K y de \mathbf{g} , tanto con priori conjugada como con una versión no conjugada. En la Sección 3.4 se efectúa un estudio de sensibilidad del mejor diseño *Fieller* para una priori conjugada y pérdida cuadrática, variando los horizontes, tamaños de región experimental y parámetros de la priori. En dicho estudio se muestra el comportamiento de los valores de K^* y \mathbf{g}^* , y el grado de mejora respecto de *certainty equivalence* del mejor diseño *Fieller*, $d_F(K^*, \mathbf{g}^*)$.

3.3 Diseño *Fieller* para modelo lineal.

Para ilustrar el uso de los diseños *Fieller*, se explora su utilización en el modelo lineal normal con $f(x; \mathbf{b}) = \mathbf{b}_0 + \mathbf{b}_1 x$. Primero se considera el caso conjugado en el que $\mathbf{p}_0(\mathbf{b})$ es $N(\mathbf{m}, \mathbf{s}^2 M^{-1})$. La superficie es estimada a través de $\hat{y}_n(x) = E_n[y(x)] = E_n[\mathbf{b}_0] + E_n[\mathbf{b}_1]x$ y usando la notación del Apéndice I, se tiene que $U_{n+1}(K)$ y $V_{n+1}(K)$ son las raíces de

$$z'(M + Z'_n Z_n)^{-1} (M\mathbf{m} + Z'_n Y_n) \pm K\mathbf{s} \sqrt{z'(M + Z'_n Z_n)^{-1} z} = T,$$

con $z' = (1, x)$. Para los detalles sobre la computación de $U_{n+1}(K)$ y $V_{n+1}(K)$ en el caso particular en el que \mathbf{b}_0 y \mathbf{b}_1 son independientes a priori ver el Apéndice 3.A.

En la primera columna de la Figura 3.3 se muestra el comportamiento de las superficies de contorno en función de K y \mathbf{g} para las tres $r_i(K, \mathbf{g})$ descritas en el Capítulo 1, con $N = 25$, $S = 100$ y $T = 0$, todo para esta primera priori conjugada. El gráfico ha sido construido a partir de los valores que adoptan $r_i(K, \mathbf{g})$ en una rejilla igualmente espaciada de 11×25 valores de (K, \mathbf{g}) . En esta primera columna la priori es $\mathbf{b}_0 \sim N(0, \mathbf{s}_{b_0}^2 = .4^2)$, $\mathbf{b}_1 \sim N(1, \mathbf{s}_{b_1}^2 = 2.^2)$, \mathbf{b}_1 independiente de \mathbf{b}_0 y $\mathbf{s}^2 = .2^2$. Como se puede apreciar los tres gráficos de contorno correspondientes a esta priori tienen un comportamiento similar para las tres pérdidas.

Para esta priori conjugada, $\hat{r}_2(d_{ce}) = \hat{r}_2(K = 0, \mathbf{g} = 0) = 2.40531$, con un error estándar de .00072, un valor típico sobre el rango de (K, \mathbf{g}) estudiado. Observar que los contornos para esta priori conjugada tienen dos mínimos locales, con $(\hat{K}^{2*}, \hat{\mathbf{g}}^{2*}) = (.62, .12)$ y $\hat{r}_2(\hat{K}^{2*}, \hat{\mathbf{g}}^{2*}) = 2.17600$, con un error estándar de .00070. Se conjetura que la K que separa ambos extremos corresponde a $K' = |\mathbf{m}_{b_1}| / \mathbf{s}_{b_1}$, que es la menor de las K que hace a $V_1(K)$ ó $U_1(K)$ igual a $+\infty$ ó $-\infty$, como se muestra en el Corolario 3.1 del Apéndice 3.A. Estos comentarios también son aplicables a las otras dos pérdidas, r_1 y r_3 .

El hecho de que \mathbf{g}^{I*} es positivo indica que el mejor diseño *Fieller* experimentará lejos de la elección *certainty equivalence* hacia el lado más ancho del intervalo $(V'_{n+1}(K), U'_{n+1}(K))$,

donde el conocimiento sobre \hat{y}_n es menos preciso. Como se podrá comprobar en la Sección 3.4 dedicada al análisis de sensibilidad del diseño *Fieller*, este hecho está influenciado principalmente por el valor de $\mathbf{s}_{b_1}^2$. Para la misma \mathbf{s}^2 , si el valor de $\mathbf{s}_{b_1}^2$ hubiera sido menor que 1.4, la \mathbf{g}^{i*} hubiese pasado a ser negativa.

El segundo caso que se considera es en el que $y_n | x_n, \mathbf{b} \sim N(\mathbf{b}_1(x - \mathbf{q}), \mathbf{s}^2)$ con $\mathbf{q} \sim U(rtl, rtu)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1}, \mathbf{s}_{b_1}^2)$ y \mathbf{b}_1 independiente de \mathbf{q} . Aquí la superficie es estimada por $\hat{y}_n(x) = \mathbf{b}_{1n}^{pmo}(x - \mathbf{q}_n^{pmo}) + T$, y para computar $U_{n+1}(K)$ y $V_{n+1}(K)$ se estima $s.d.(\hat{y}_n(x))$ como la raíz cuadrada de $Var(\mathbf{b}_{1n}^{pmo}(x - \mathbf{q}_n^{pmo}) | \mathbf{q}_n^{pmo} \in (rtl, rtu))$ dada en el Apéndice II. Recordar que $(\mathbf{q}_n^{pmo}, \mathbf{b}_{1n}^{pmo})$ es la moda a posteriori de $\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1)$.

Es importante destacar que si \mathbf{q}_n^{pmo} coincidiera con uno de los extremos de la región experimental C , por ejemplo, si $\mathbf{q}_n^{pmo} = rtl$, $Var(\mathbf{b}_{1n}^{pmo}(x - \mathbf{q}_n^{pmo}) | \mathbf{q}_n^{pmo} = rtl)$ sería igual a $(x_0 - rtl)^2 Var(\mathbf{b}_{1n}^{pmo} | \mathbf{q}_n^{pmo} = rtl)$ y $U_{n+1}(K)$ y $V_{n+1}(K)$ valdrían rtl para cualquier K , con lo que el intervalo *Fieller* degenera en un punto. En ese caso lo que hacemos es calcular el intervalo *Fieller* sin la restricción de que el estimador de $\mathbf{q} \in [rtl, rtu]$, para luego trasladar tanto el intervalo *Fieller* como el estimador de \mathbf{q} al extremo del intervalo $[rtl, rtu]$ que corresponda. Otra alternativa sería considerar la desviación típica de la predicción de la respuesta, $s.d.(\hat{y}_n(x) + \mathbf{e})$, en vez de $s.d.(\hat{y}_n(x))$, lo que evitaría la degeneración del intervalo *Fieller* cuando \mathbf{q}_n^{pmo} coincidiera con un extremo.

Los gráficos de contorno de las $r_i(K, \mathbf{g})$ estimadas de la segunda columna de la Figura 3.3 son para la priori no conjugada $\mathbf{q} \sim U(-.5, .5)$, $\mathbf{b}_1 \sim N(1., \mathbf{s}_{b_1}^2 = .2^2)$, \mathbf{b}_1 independiente de \mathbf{q} y $\mathbf{s}^2 = .4^2$. Al igual que ocurría para el caso conjugado, los gráficos de contorno para las tres pérdidas son muy similares y, por tanto, nos centramos en el caso de la pérdida cuadrática. Para esta priori, $\hat{r}_2(d_{ce}) = \hat{r}_2(K=0, \mathbf{g}=0) = 4.54111$, con un error estándar de .00020, $(K^{2*}, \mathbf{g}^{2*})$ es estimado a ser $(\hat{K}^{2*}, \hat{\mathbf{g}}^{2*}) = (2.75, .125)$ con $\hat{r}_2(\hat{K}^{2*}, \hat{\mathbf{g}}^{2*}) = 4.48778$ y un error estándar de .00018. Destaca para esta priori que se puede aumentar $|\mathbf{g}|$ sin decrecer significativamente r_i

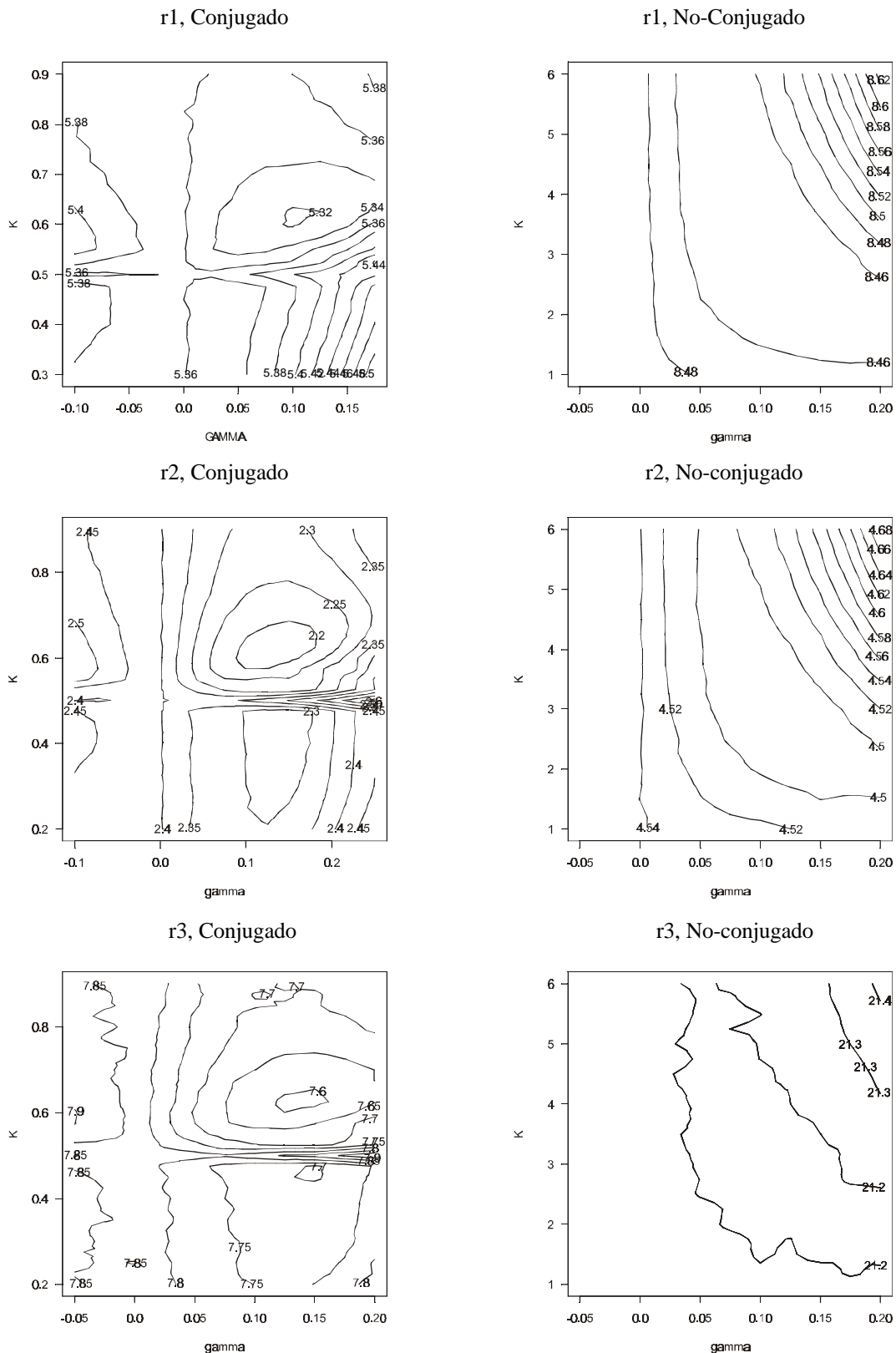


Figura 3.3. Gráfico de contorno de $\hat{r}_i(K, \mathbf{g})$, $i=1,2,3$, con $N=25$, $S=100$, $T=0$ y $C=[-1,1]$. La primera columna es para la priori $\mathbf{b}_0 \sim N(\mathbf{m}_{b_0} = 0., \mathbf{s}_{b_0}^2 = .4^2)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1} = 1., \mathbf{s}_{b_1}^2 = 2.^2)$, \mathbf{b}_0 independiente de \mathbf{b}_1 y $\mathbf{s}^2 = .2^2$. La segunda columna es para $\mathbf{q} \sim U(-.5, .5)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1} = 1., \mathbf{s}_{b_1}^2 = 2.^2)$, \mathbf{b}_1 independiente de \mathbf{q} y $\mathbf{s}^2 = .4^2$.

al disminuir de forma apropiada K . El \mathbf{g}^{t*} es positivo lo que indica que para estos parámetros de la priori el mejor diseño *Fieller* experimentará lejos de *certainty equivalence* hacia el lado más ancho del intervalo $(V'_{n+1}(K), U'_{n+1}(K))$.

3.4 Análisis de sensibilidad.

En la Sección 3.3 hemos presentado el comportamiento de $r(d)$ en función de K y \mathbf{g} para una priori concreta. Aquí en cambio estudiamos el comportamiento de K^* y \mathbf{g}^* , así como la mejora relativa de $d_F(K^*, \mathbf{g}^*)$, en función de la priori, N y tamaño de C , todo para la pérdida cuadrática $r_2(K, \mathbf{g})$.

Para evaluar la mejora relativa del mejor diseño *Fieller*, $d_F(K^*, \mathbf{g}^*)$, con respecto a cualquier $d_F(K, \mathbf{g})$ se define $rr(K, \mathbf{g})$ como la diferencia relativa entre la pérdida esperada para $d_F(K^*, \mathbf{g}^*)$ y la pérdida esperada para $d_F(K, \mathbf{g})$,

$$rr(K, \mathbf{g}) = \frac{r_2(K, \mathbf{g}) - r_2(K^*, \mathbf{g}^*)}{r_2(K^*, \mathbf{g}^*)}.$$

En particular, notar que $rr(0, \mathbf{g})$ es la mejora relativa de $d_F(K^*, \mathbf{g}^*)$ sobre el diseño *certainty equivalence*.

Se estudia el comportamiento del $rr(K, \mathbf{g})$ para el modelo lineal normal con $f(x; \mathbf{b}) = \mathbf{b}_0 + \mathbf{b}_1 x$ y priori conjugada, con $\mathbf{p}_0(\mathbf{b})$ tal que $\mathbf{b}_0 \sim N(0, \mathbf{s}_{b_0}^2 = .5^2)$, $\mathbf{b}_1 \sim N(1, \mathbf{s}_{b_1}^2)$ con \mathbf{b}_1 independiente de \mathbf{b}_0 , $T=0$ y $C=[-R, R]$. Para evaluar $r_2(K, \mathbf{g})$ se simulan 3.000.000 de $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1)$ según $\mathbf{p}_0(\mathbf{b})$, y para cada \mathbf{b} se ejecuta el diseño $d_F(K, \mathbf{g})$ sobre las $y_i(x_i^F)$'s simuladas usando la distribución de $(y_i | x_i^F, \mathbf{b})$ y promediando las pérdidas observadas. Como la desviación estándar estimada de $\hat{r}_2(K, \mathbf{g})$ es del orden de .0004, el valor

(K^*, \mathbf{g}^*) que minimiza $r_2(K, \mathbf{g})$ se puede obtener fácilmente de forma muy precisa a través del procedimiento dado en Nelder y Mead (1965).

La Figura 3.4 muestra como K^* , \mathbf{g}^* y $rr(0, \mathbf{g})$ dependen de \mathbf{s}_{b_1} , de \mathbf{s} , del horizonte N y del tamaño de la región experimental R . Cuando $N = 20$ y $R = 10$, K^* y \mathbf{g}^* dependen de \mathbf{s} y \mathbf{s}_{b_1} de una forma altamente no lineal. Para la mayoría de las combinaciones de $(\mathbf{s}_{b_1}, \mathbf{s})$ utilizadas, \mathbf{g}^* es negativo y esto indica que el mejor diseño *Fieller* experimenta en el subintervalo más estrecho de $[V'_{n+1}(K), U'_{n+1}(K)]$, perturbando la elección *certainty equivalence* en la dirección en la que se encuentran la mayoría de las observaciones ya tomadas. Por otro lado, K^* es bastante insensible a las variaciones de N , mientras que \mathbf{g}^* tiende hacia cero cuando N tiende a infinito. Así, cuanto más grande es N más se acerca el diseño *Fieller* al diseño *certainty equivalence*. Es importante remarcar que K^* es bastante sensible a los cambios en el tamaño de la región experimental R . Cuanto mayor es R mayor es K^* y, por tanto, más lejos del *certainty equivalence* puede experimentar el diseño *Fieller*.

La Figura 3.4 también prueba que $rr(0, \mathbf{g})$ aumenta cuando aumenta \mathbf{s}_{b_1} o disminuye \mathbf{s} . Cuanto menos se sabe de \mathbf{b}_1 y menor es la variabilidad de y sobre su media mayor será la mejora sobre *certainty equivalence*. Por otro lado, la mejora relativa sobre *certainty equivalence* disminuye cuando aumenta N , hecho que es consistente con que para modelos lineales simples los diseños *certainty equivalence* Bayes son asintóticamente óptimos, Chen y Hu (1998). Notar también que $rr(0, \mathbf{g})$ crece bastante para R grande, debido al comportamiento extremadamente pobre del diseño *certainty equivalence* en tales circunstancias. En general, para todos los ejemplos que se han probado se encontró un comportamiento similar de K^* , \mathbf{g}^* y $rr(0, \mathbf{g})$ en función de \mathbf{s} , \mathbf{s}_{b_1} , R y N .

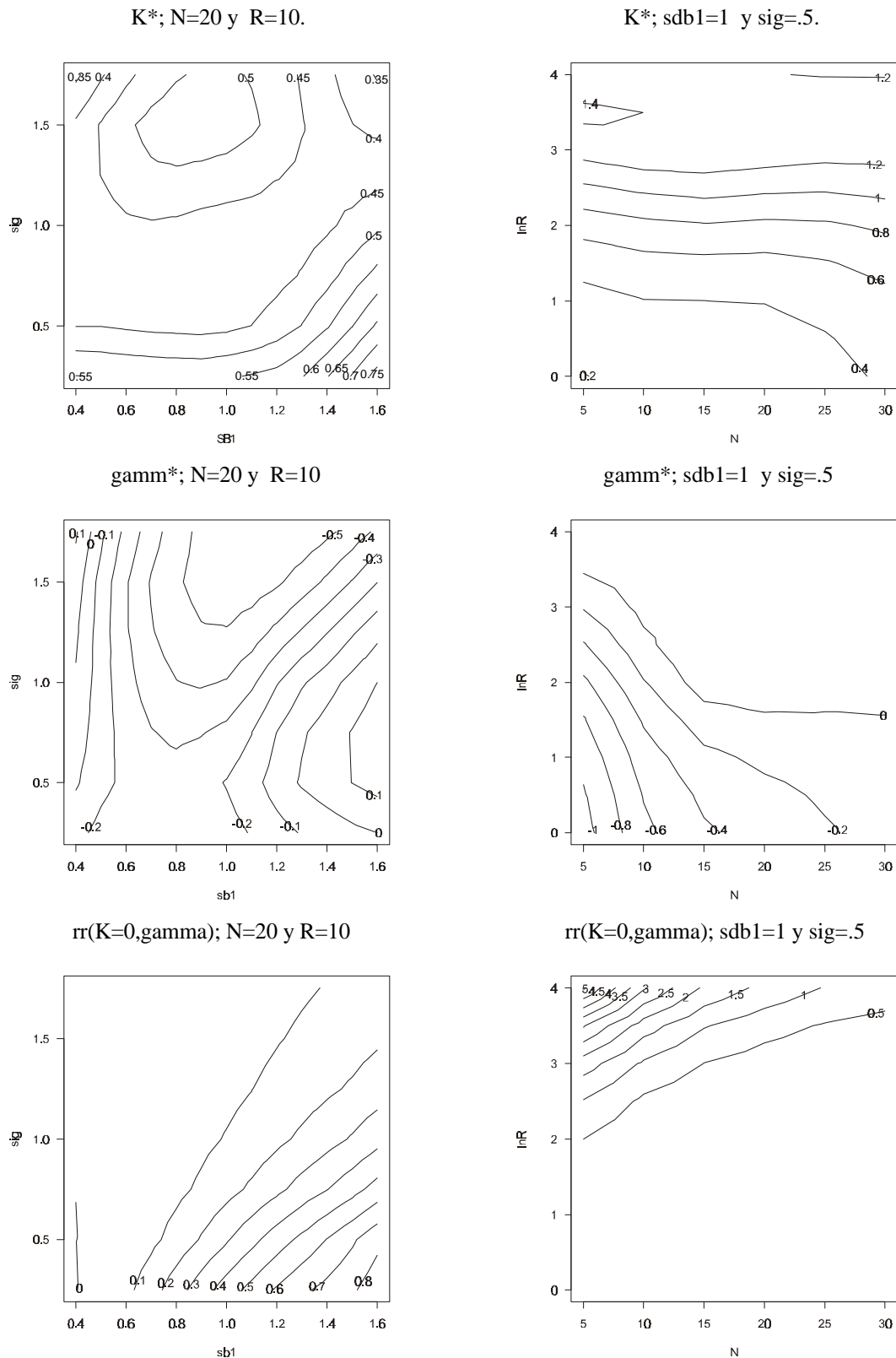


Figura 3.4. Gráfico de contorno de K^* , g^* y de la mejora relativa de $d_F(K^*, g^*)$ sobre c.e., $rr(0, g)$, como una función de σ , s_{b_1} , N y $\ln(R)$. $(y|x, \mathbf{b}) \sim N(\mathbf{b}_0 + \mathbf{b}_1 x, \mathbf{S}^2)$ con $\mathbf{b}_0 \sim N(\mathbf{m}_{b_0} = 0., \mathbf{S}_{b_0}^2 = .5^2)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1} = 1., \mathbf{S}_{b_1}^2)$, \mathbf{b}_1 independiente de \mathbf{b}_0 , $C = [-R, R]$ y $T = 0$.

3.5 Extensiones y comentarios.

Los diseños *Fieller* introducen una acción de sondeo entorno a la elección *certainty equivalence* Bayesiana que le permite mejorar los estimadores de los parámetros \mathbf{b} en los estadios intermedios y así conseguir obtener un mejor control futuro, a la vez que aprovechan el que asintóticamente los diseños *certainty equivalence* son óptimos. En estos diseños la acción de sondeo se realiza sobre unos intervalos que son aproximaciones truncadas de los intervalos *Fieller* Bayesianos, cuyo cálculo es relativamente fácil incluso cuando la distribución a posteriori sobre \mathbf{q} no se puede obtener de forma sencilla, aunque precisan de la estimación repetida de $s.d.(\hat{y}_n(x))$.

Si la distribución a posteriori para \mathbf{q} se puede hallar de forma sencilla, es factible aplicar esta misma idea de los diseños *Fieller* usando conjuntos crebles a posteriori de \mathbf{q} , tales como intervalos centrales a posteriori o regiones *HPD*, tal como se mostrará en el Capítulo 4.

Cuando $x \in R^k$, la solución a $f(x; \mathbf{b}) = T$ será una superficie y no un número real, lo que hace que los diseños *Fieller* no puedan ser adaptados de forma directa a problemas con múltiples variables control. Una solución al problema con $x \in R^k$ es propuesta en el Capítulo 6.

Las mejoras relativas de *Fieller* sobre *certainty equivalence* dependerán de la información a priori, del horizonte y de la región experimental, y así el valor de dicha mejora estará en función del contexto particular donde se aplica. En algunos ejemplos podría ser útil adaptar nuevamente estos diseños recomputando $(\hat{K}_m^*, \hat{\mathbf{g}}_m^*)$ en todos o en algunos m estadios intermedios, basándose en la distribución a posteriori de \mathbf{b} en ese estadio. No obstante, esta mejora tiene un gran coste computacional y no siempre será sustancial. Por ejemplo, la obtención de $(\hat{K}_2^*, \hat{\mathbf{g}}_2^*)$ obligaría a realizar todo el proceso de simulación (para las N observaciones) y minimización de la estimación de la pérdida esperada en función de K y \mathbf{g} para calcular el $(\hat{K}_1^*, \hat{\mathbf{g}}_1^*)$, que coincidirá con el

$(\hat{K}^*, \hat{\mathbf{g}}^*)$ dados en este capítulo, para nuevamente realizar todo el proceso de simulación (hallando la primera observación haciendo uso del $(\hat{K}_1^*, \hat{\mathbf{g}}_1^*)$) y minimización de la estimación de la pérdida esperada en función de K y \mathbf{g} para la obtención de $(\hat{K}_2^*, \hat{\mathbf{g}}_2^*)$.

Como ya se ha indicado, $U_{n+1}(K)$ y $V_{n+1}(K)$ son las raíces de $\hat{y}_n(x) \pm K \hat{s.d.}(\hat{y}_n(x)) = T$ o, de forma equivalente, las raíces de $(\hat{y}_n(x) - T)^2 - K^2 \hat{Var}(\hat{y}_n(x)) = 0$. Se puede comprobar de forma sencilla que si $\hat{y}_n(x) = E_n[y(x)]$ el x que minimiza $(\hat{y}_n(x) - T)^2 - K^2 \hat{Var}(\hat{y}_n(x))$ cuando $K^2 = -1$, da lugar al diseño *miope* y cuando $K = 0$ al diseño *certainty equivalence*. Además, en el modelo lineal esta última expresión es una forma cuadrática y, por tanto, la minimización de dicha expresión coincidirá con el centro de dicha forma cuadrática. El hecho de que en una misma expresión y con un único parámetro se pueda obtener tanto el diseño *certainty equivalence* como el *miope* será aprovechado en el Capítulo 6 para la construcción de los diseños *híbridos*.

Apéndice 3.A Cálculo de intervalos *Fieller* para el modelo lineal.

3.A.1 Priori conjugada con b_0 independiente de b_1 .

Proposición 3.1 Sean $Y_n | X_n, \mathbf{b} \sim N_n(Z_n \mathbf{b} + 1_n T, \mathbf{S}^2 I_n)$, $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1) \sim N_2((\mathbf{m}_{b_0}, \mathbf{m}_{b_1}), \mathbf{S}^2 M^{-1})$ con M una matriz diagonal (m_{00}, m_{11}) . Los valores de $V_{n+1}(K)$ y $U_{n+1}(K)$ asociados al diseño *Fieller* $d_F(K, \mathbf{g})$ se obtienen resolviendo la ecuación

$$x^2(b_n^2 - D_n(m_{00} + n)) + 2x(a_n b_n + D_n \sum x_i) + a_n^2 - D_n(m_{11} + \sum x_i^2) = 0,$$

donde $L_n = |M + Z_n' Z_n|$, $a_n = L_n E_n[\mathbf{b}_0]$, $b_n = L_n E_n[\mathbf{b}_1]$ y $D_n = K^2 \mathbf{S}^2 L_n$.

Demostración.

La solución de $\hat{y}_n(x) \pm K \hat{s}.d.(\hat{y}_n(x)) = 0$ coincide con la solución de $(\hat{y}_n(x))^2 - K^2 \hat{Var}(\hat{y}_n(x)) = 0$. Usando los resultados distribucionales dados en el Apéndice I para dicha priori, se llega a la ecuación anterior. ■

Notar que al resolver la ecuación de segundo grado de la proposición anterior se pueden obtener dos, una o ninguna solución. En el caso de una o ninguna solución, $V_{n+1}(K)$ ó $U_{n+1}(K)$ valdrá ∞ . Cuando existan dos soluciones, habrá que comprobar si ambas contienen el punto $x_{n+1} = -a_n/b_n$, o si se encuentran al mismo lado de $x_{n+1} = -a_n/b_n$. En este segundo caso se considera que uno de los extremos del intervalo es ∞ y el otro la solución más próxima a x_{n+1} . El siguiente corolario resuelve el problema para el estadio inicial, antes de empezar a recoger datos.

Corolario 3.1 Sean $Y_n | X_n, \mathbf{b} \sim N_n(Z_n \mathbf{b} + 1_n T, \mathbf{S}^2 I_n)$, $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1) \sim N_2((\mathbf{m}_{b_0}, \mathbf{m}_{b_1}), \mathbf{S}^2 M^{-1})$

con M una matriz diagonal (m_{00}, m_{11}) :

a) Si $K = 0$ entonces $V_1(0) = U_1(0) = x_1 = -\mathbf{m}_{b_0} / \mathbf{m}_{b_1}$.

b) Si $0 < K^2 < \frac{\mathbf{m}_{b_1}^2}{\mathbf{s}_{b_1}^2}$ entonces $V_1(K)$ y $U_1(K)$ tienen valores finitos y contienen a

$$x_1 = -\mathbf{m}_{b_0} / \mathbf{m}_{b_1}.$$

c) Si $\frac{\mathbf{m}_{b_1}^2}{\mathbf{s}_{b_1}^2} \leq K^2 \leq \frac{\mathbf{m}_{b_1}^2}{\mathbf{s}_{b_1}^2} + \frac{\mathbf{m}_{b_0}^2}{\mathbf{s}_{b_0}^2}$ entonces ó $V_1(K)$ ó $U_1(K)$ es infinito.

d) Si $\frac{\mathbf{m}_{b_1}^2}{\mathbf{s}_{b_1}^2} + \frac{\mathbf{m}_{b_0}^2}{\mathbf{s}_{b_0}^2} < K^2$ entonces $V_1(K)$ y $U_1(K)$ son infinitos.

Demostración.

Como caso particular de la Proposición 3.1, $V_1(K)$ y $U_1(K)$ son la solución de la ecuación

$$x^2(m_{00}m_{11}\mathbf{m}_{b_1}^2 - K^2\mathbf{s}^2m_{00}) + 2x m_{00}m_{11}\mathbf{m}_{b_0}\mathbf{m}_{b_1} + m_{00}m_{11}\mathbf{m}_{b_0}^2 - K^2\mathbf{s}^2m_{11} = 0.$$

Cuando $\mathbf{m}_{b_0} = 0$ dicha ecuación se reduce a

$$x^2(m_{00}m_{11}\mathbf{m}_{b_1}^2 - K^2\mathbf{s}^2m_{00}) - K^2\mathbf{s}^2m_{11} = 0,$$

de donde se sigue que si $(m_{00}m_{11}\mathbf{m}_{b_1}^2 - K^2\mathbf{s}^2m_{00}) \leq 0$ entonces $K^2 \geq \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2$ y, por tanto, no existe solución, con lo que $V_1(K) = -\infty$ y $U_1(K) = +\infty$. Cuando $K^2 < \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2$ se obtienen las dos soluciones finitas, que además son simétricas respecto a $x_1 = -\mathbf{m}_{b_0} / \mathbf{m}_{b_1} = 0$.

Supongamos, sin pérdida de generalidad, que $\mathbf{m}_{b_0} > 0$. Cuando $m_{00}m_{11}\mathbf{m}_{b_1}^2 - K^2\mathbf{s}^2m_{00} = 0$ se tiene que $K^2 = \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2$, y la solución de la ecuación en función de K es única e igual a

$$x = \frac{1}{2} \left(-\frac{\mathbf{m}_{b_0}}{\mathbf{m}_{b_1}} + \frac{m_{11}\mathbf{m}_{b_1}}{m_{00}\mathbf{m}_{b_0}} \right) = \frac{1}{2} \left(x_1 + \frac{\mathbf{s}_{b_0}^2 \mathbf{m}_{b_1}}{\mathbf{s}_{b_1}^2 \mathbf{m}_{b_0}} \right).$$

El intervalo será de la forma $(-\infty, x)$ ó $(x, +\infty)$ dependiendo del signo de \mathbf{m}_{b_1} . Por otro lado, cuando $m_{00}m_{11}\mathbf{m}_{b_1}^2 - K^2\mathbf{s}^2m_{00} \neq 0$ la solución de la ecuación en función de K tendrá la forma

$$x = \frac{-m_{00}m_{11}\mathbf{m}_{b_0}\mathbf{m}_{b_1} \pm \sqrt{D}}{m_{00}m_{11}\mathbf{m}_{b_1}^2 - K^2\mathbf{s}^2m_{00}},$$

con $D = K^2 \mathbf{s}^2 m_{00} m_{11} (m_{00} \mathbf{m}_{b_0}^2 + m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2)$, de forma que:

- a) Si $D = 0$ se pueden dar dos casos. Cuando $K = 0$ entonces $x = -\mathbf{m}_{b_0} / \mathbf{m}_{b_1}$ y, por tanto, $U_1 = V_1 = x$, y cuando $K^2 = \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2 + \mathbf{m}_{b_0}^2 / \mathbf{s}_{b_0}^2$ entonces $x = (\mathbf{s}_{b_0}^2 \mathbf{m}_{b_1}) / (\mathbf{s}_{b_1}^2 \mathbf{m}_{b_0})$ y el intervalo es de la forma $(-\infty, x)$ ó $(x, +\infty)$.
- b) Si $D > 0$ entonces $K^2 < \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2 + \mathbf{m}_{b_0}^2 / \mathbf{s}_{b_0}^2$ y, consecuentemente, existen dos soluciones, aunque falta comprobar si la opción *certainty equivalence* está entre las dos soluciones, es decir, si

$$\frac{-m_{00} m_{11} \mathbf{m}_{b_0} \mathbf{m}_{b_1} - \sqrt{D}}{m_{00} m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2 m_{00}} < x_1 = -\frac{\mathbf{m}_{b_0}}{\mathbf{m}_{b_1}} < \frac{-m_{00} m_{11} \mathbf{m}_{b_0} \mathbf{m}_{b_1} + \sqrt{D}}{m_{00} m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2 m_{00}}.$$

1. Si $m_{00} m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2 m_{00} > 0$ entonces $K^2 < \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2$ y, por lo tanto, se tiene por un lado de la desigualdad, el izquierdo si $\mathbf{m}_{b_1} > 0$ o el derecho si $\mathbf{m}_{b_1} < 0$, que $-|\mathbf{m}_{b_1}| \sqrt{D} < 0 < K^2 \mathbf{s}^2 \mathbf{m}_{b_0} m_{00}$, que se verifica siempre. Por el otro lado de la desigualdad se sigue que $0 < (m_{00} \mathbf{m}_{b_0}^2 + m_{11} \mathbf{m}_{b_1}^2)(m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2)$, que también se verifica, lo que hace que

$$(V_1, U_1) = \left(\frac{-m_{00} m_{11} \mathbf{m}_{b_0} \mathbf{m}_{b_1} - \sqrt{D}}{m_{00} m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2 m_{00}}, \frac{-m_{00} m_{11} \mathbf{m}_{b_0} \mathbf{m}_{b_1} + \sqrt{D}}{m_{00} m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2 m_{00}} \right).$$

2. Si $m_{00} m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2 m_{00} < 0$ entonces $K^2 > \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2$ y, análogamente, se tiene por un lado de la desigualdad, el izquierdo si $\mathbf{m}_{b_1} > 0$ o el derecho si $\mathbf{m}_{b_1} < 0$, que $-|\mathbf{m}_{b_1}| \sqrt{D} > K^2 \mathbf{s}^2 \mathbf{m}_{b_0} m_{00}$, que no se verifica, lo que hace que

$$(V_1, U_1) = \left(-\infty, \frac{-m_{00} m_{11} \mathbf{m}_{b_0} \mathbf{m}_{b_1} - \sqrt{D}}{m_{00} m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2 m_{00}} \right) \quad \text{ó} \quad (V_1, U_1) = \left(\frac{-m_{00} m_{11} \mathbf{m}_{b_0} \mathbf{m}_{b_1} + \sqrt{D}}{m_{00} m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2 m_{00}}, +\infty \right),$$

según si $\mathbf{m}_{b_1} > 0$ o $\mathbf{m}_{b_1} < 0$, respectivamente.

- c) Si $D < 0$ entonces $K^2 = \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2 + \mathbf{m}_{b_0}^2 / \mathbf{s}_{b_0}^2$ y no existe solución, y el intervalo es de la forma $(-\infty, +\infty)$. ■

A menudo el x_1^{ce} será igual a x_1 ya que al especificar la priori lo corriente será que $x_1 \in C$. Obsérvese que cuando $K^2 = \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2$, el x obtenido como extremo finito del intervalo $(V_1(K), U_1(K))$, es el punto medio entre el x_1^{ce} y el extremo finito de dicho intervalo para $K^2 = \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2 + \mathbf{m}_{b_0}^2 / \mathbf{s}_{b_0}^2$.

El siguiente corolario resuelve el intervalo *Fieller* para el estadio 1, cuando ya se ha observado en un punto, en el caso particular en que $\mathbf{m}_{b_0} = 0$ y la región C está formada por un intervalo simétrico respecto a cero. Esto corresponde a la situación encontrada en el ejemplo de las Secciones 3.3 y 3.4.

Corolario 3.2 Sean $Y_n | X_n, \mathbf{b} \sim N_n(Z_n \mathbf{b} + 1_n T, \mathbf{s}^2 I_n)$, $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1) \sim N_2((\mathbf{m}_{b_0}, \mathbf{m}_{b_1}), \mathbf{s}^2 M^{-1})$ con M una matriz diagonal (m_{00}, m_{11}) , $\mathbf{m}_{b_0} = 0$ y la región C formada por un intervalo simétrico respecto a cero. Se verifica que $x_1^F(K, \mathbf{g}) = x_1^{ce} = 0$ y además

- a) Si $K = 0$ entonces $V_2(0) = U_2(0) = x_2 = -\frac{y_1}{(1 + m_{00})\mathbf{m}_{b_1}}$.
- b) Si $0 < K^2 < \frac{\mathbf{m}_{b_1}^2}{\mathbf{s}_{b_1}^2}$ entonces $V_2(K)$ y $U_2(K)$ tienen valores finitos y encierran a x_2 .
- c) Si $\frac{\mathbf{m}_{b_1}^2}{\mathbf{s}_{b_1}^2} \leq K^2 \leq \frac{\mathbf{m}_{b_1}^2}{\mathbf{s}_{b_1}^2} + \frac{y_1^2}{\mathbf{s}^2(1 + m_{00})}$ entonces ó $V_2(K)$ ó $U_2(K)$ es infinito.
- d) Si $\frac{\mathbf{m}_{b_1}^2}{\mathbf{s}_{b_1}^2} + \frac{y_1^2}{\mathbf{s}^2(1 + m_{00})} < K^2$ entonces $V_2(K)$ y $U_2(K)$ son infinitos.

Demostración.

Haciendo uso del Corolario 3.1, si $K^2 \geq \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2$, $U_1(K)$ y $V_1(K)$ serán iguales a infinito y el intervalo *Fieller* truncado $(V_1'(K), U_1'(K))$ coincide con C . Al ser C simétrico respecto a $x_1 = x_1^{ce} = -\mathbf{m}_{b_0} / \mathbf{m}_{b_1} = 0$, el diseño *Fieller* observará en $x_1^F(K, \mathbf{g}) = x_1^{ce} = 0$. Si $K^2 < \mathbf{m}_{b_1}^2 / \mathbf{s}_{b_1}^2$, el intervalo $(V_1(K), U_1(K))$ es simétrico respecto a cero, y también se obtiene $x_1^F(K, \mathbf{g}) = x_1^{ce} = 0$.

Si se ha observado y_1 en $x_1 = 0$, sustituyendo en la ecuación de la Proposición 3.1 cuando $n=1$, se llega a la ecuación

$$x^2(1+m_{00})^2 m_{11}(m_{11} \mathbf{m}_{b_1}^2 - K^2 \mathbf{s}^2) + 2x(1+m_{00})m_{11} \mathbf{m}_{b_1} y_1 + m_{11}^2 y_1^2 - K^2 \mathbf{s}^2 m_{11}^2 (1+m_{00}) = 0.$$

El resto de la demostración es equivalente a la del Corolario 3.1. ■

Notar que cuando $\mathbf{m}_{b_0} = 0$, tanto para $(V_1(K), U_1(K))$ como para $(V_2(K), U_2(K))$, el primer valor de K que hace que al menos uno de los dos extremos sea infinito es $K = |\mathbf{m}_{b_1}| / \mathbf{s}_{b_1}$. Si $\mathbf{m}_{b_0} \neq 0$ esto sólo se verifica para $(V_1(K), U_1(K))$.

3.A.2 Priori no conjugada.

Proposición 3.2 Sean $(y_n | x_n, \mathbf{q}, \mathbf{b}_1) \sim N(\mathbf{b}_1(x_n - \mathbf{q}) + T, \mathbf{s}^2)$ con $\mathbf{q} \sim U(\text{rtl}, \text{rtu})$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1}, \mathbf{s}_{b_1}^2)$, independientes y con \mathbf{s} conocida. Entonces el intervalo $(V_{n+1}(K), U_{n+1}(K))$ asociado al diseño Fieller $d_F(K, \mathbf{g})$ se obtiene a partir de la resolución de

$$x^2(b_n^2 - D_n n) + 2x(-b_n^2 \mathbf{q}_n^{pmo} + D_n \sum x_i) + (b_n \mathbf{q}_n^{pmo})^2 - D_n \sum x_i^2 - L_n = 0,$$

con $D_n = K^2 \mathbf{s}^2 \mathbf{s}_{b_1}^2 w_{1n}$, $L_n = K^2 \mathbf{s}^2 (2w_{1n} + w_{2n}) \mathbf{s}^2$, denotando por $w_{1n} = n \mathbf{s}_{b_1}^2 \sum (x_i - \bar{x}_n)^2$, $w_{2n} = n \mathbf{s}^2$ y $b_n = n \mathbf{s}_{b_1}^2 \sum (x_i - \bar{x}_n)(y_i - \bar{y}_n) + n \mathbf{s}^2 \mathbf{m}_{b_1}$.

Demostración.

Siguiendo el procedimiento indicado en la Proposición 3.1 para resolver $(\hat{y}_n(x))^2 - K^2 \text{Var}(\hat{y}_n(x)) = 0$, y utilizando los estimadores de $y_n(x)$ y de $\text{Var}(\hat{y}_n(x))$ dados en la Sección 1 del Apéndice II, se obtiene la ecuación expuesta en esta proposición. En el caso en que \mathbf{q}_n^{pmo} sea uno de los extremos del intervalo $[\text{rtl}, \text{rtu}]$ la $\text{Var}(\hat{y}_n(x))$, según los estimadores

del Apéndice II para esta situación, será sustituida por la que se obtiene cuando se reemplaza rtl ó rtu en la expresión para $\mathbf{q}_n^{pmo} \in (rtl, rtu)$. ■

Los dos corolarios siguientes muestran la forma del intervalo $(V_n(K), U_n(K))$ para esta priori en el primero y segundo estadio.

Corolario 3.3 Sean $(y_n | x_n, \mathbf{q}, \mathbf{b}_1) \sim N(\mathbf{b}_1(x_n - \mathbf{q}) + T, \mathbf{S}^2)$ con $\mathbf{q} \sim U(rtl, rtu)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1}, \mathbf{S}_{b_1}^2)$, independientes y con \mathbf{s} conocida. Entonces el intervalo $(V_0(K), U_0(K))$ asociado al diseño Fieller, $d_F(K, \mathbf{g})$, se reduce al único punto $x_1^F(K, \mathbf{g}) = x_1^{ce} = (rtl + rtu)/2$, para todo K y \mathbf{g}

Demostración.

Al definir $\mathbf{q}_0^{pmo} = (rtl + rtu)/2$ como el punto medio del intervalo soporte de una distribución uniforme, y al tener que $\mathbf{b}_{10}^{pmo} = \mathbf{m}_{b_1}$, se obtiene que la desviación tipo de $\hat{y}(x; \mathbf{q}, \mathbf{b}_1)$ es igual a cero y, por tanto, $V_0(K)$ y $U_0(K)$ son la solución a

$$\mathbf{b}_{10}^{pmo} (x - \mathbf{q}_0^{pmo}) = 0. \quad \blacksquare$$

Corolario 3.4 Sean $(y_n | x_n, \mathbf{q}, \mathbf{b}_1) \sim N(\mathbf{b}_1(x_n - \mathbf{q}) + T, \mathbf{S}^2)$ con $\mathbf{q} \sim U(rtl, rtu)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1}, \mathbf{S}_{b_1}^2)$, independientes y con \mathbf{s} conocida. Entonces el intervalo $(V_1(K), U_1(K))$ cuando se ha observado y_1 en $x_1^F = (rtl + rtu)/2$ es:

$$a) \quad (V_1(K), U_1(K)) = \left(x_1^F - \frac{y_1}{\mathbf{m}_{b_1}} \pm \frac{K\mathbf{s}}{|\mathbf{m}_{b_1}|} \right) \text{ siempre que } x_1^F - y_1/\mathbf{m}_{b_1} \in [rtl, rtu],$$

$$b) \quad (V_1(K), U_1(K)) = \left(rtl \pm \frac{K\mathbf{s}}{|\mathbf{m}_{b_1}|} \right) \text{ siempre que } x_1^F - y_1/\mathbf{m}_{b_1} < rtl,$$

$$c) \quad (V_1(K), U_1(K)) = \left(rtu \pm \frac{K\mathbf{s}}{|\mathbf{m}_{b_1}|} \right) \text{ siempre que } x_1^F - y_1/\mathbf{m}_{b_1} > rtu.$$

Demostración.

Aplicando la Proposición 3.2 al caso $n=1$ y considerando por separado el caso en que $\mathbf{q}_1 = x_1^F - y_1/\mathbf{m}_{b_1}$ está dentro del intervalo $[rtl, rtu]$ del caso en que está fuera, se tiene que $V_1(K)$ y $U_1(K)$ son las soluciones de

$$x^2 - 2x\mathbf{q}_1^{pmo} + (\mathbf{q}_1^{pmo})^2 - K^2\mathbf{s}^2/\mathbf{m}_{b_1}^2 = 0. \blacksquare$$

CAPITULO 4.

DISEÑOS BASADOS EN REGIONES CREÍBLES.

Para resolver el problema de control multiperiodo, en el Capítulo 3 se introducían los diseños *Fieller* construidos sobre intervalos creíbles aproximados para la raíz \mathbf{q} a partir de las soluciones calculadas de las ecuaciones $\hat{y}_n(x) \pm K \hat{s.d.}(\hat{y}_n(x)) = T$. Dichos diseños utilizan estos intervalos de guía para decidir cuánto y hacia dónde moverse con respecto a la elección *certainty equivalence* y con ello incentivar el aprendizaje activo sobre \mathbf{q} y mejorar así el control futuro.

En el problema del *Response Surface Bandit (RSB)* definido en Ginebra y Clayton (1995), el objetivo es encontrar diseños que maximicen $w(d) = E[U(X_N^d, Y_N)]$, donde $U(X_N^d, Y_N)$ es una utilidad que valora la cercanía entre Y_N' y el máximo de $f(x; \mathbf{b})$, como pueden ser $U_1(X_N^d, Y_N) = \sum_{n=1}^N y_n(x_n^d)$. Los diseños *certainty equivalence* en este caso observan y_{n+1} en algún estimador del valor \mathbf{k} , que maximiza en x a $f(x; \mathbf{b})$, $x_{n+1}^{ce} = \hat{\mathbf{k}}_n$. Ginebra y Clayton (1995) perturban *certainty equivalence* experimentando sobre el x_{n+1} que maximiza una cota superior de la superficie prevista, $\hat{f}_n(x; \mathbf{b})$.

Cuando se selecciona x_n nos enfrentamos a dos metas en conflicto. Por un lado, necesitamos que x_n sea tal que $y_n(x_n)$ sea grande (para el *RSB*), pero por otro lado, se debe usar $(x_n, y_n(x_n))$ para aprender sobre $f(x; \mathbf{b})$ y ayudar así a mejorar la ejecución en las

últimas etapas. Este dilema es lo que se conoce en la literatura de control adaptativo como el aspecto dual de control, Astrom y Wittenmark (1995).

Las dos heurísticas comentadas en los párrafos anteriores son diferentes para ambos problemas pero tienen en común que ambas requieren la estimación de la desviación estándar de $\hat{f}_n(x; \mathbf{b})$. En este capítulo se proponen nuevas heurísticas basadas en la construcción de conjuntos creíbles para \mathbf{q} y \mathbf{k} propiamente dichos, y que pueden ser aplicadas tanto al problema de control multiperiodo como al *RSB*. En la Sección 4.1 se definirán las regiones creíbles y los distintos elementos que la componen que serán de utilidad a la hora de la construcción de los diseños propuestos en este capítulo. En las Secciones 4.2 y 4.3 se ilustra nuestra propuesta implementando diseños construidos sobre regiones a posteriori de máxima densidad sobre \mathbf{q} (regiones *HPD*), al problema de control multiperiodo con pérdida cuadrática. En las Secciones 4.4 y 4.5 se implementan diseños construidos sobre intervalos a posteriori basados en los cuantiles a posteriori de \mathbf{k} al problema *RSB*. Para ambas implementaciones se realiza un análisis de sensibilidad.

4.1 Regiones creíbles.

En esta sección se definen las dos familias de regiones creíbles que serán la base de los diseños implementados en este capítulo. La primera es la región de máxima densidad de la distribución a posteriori, que será denotada por región *HPD* (highest posteriori density) y la segunda es el intervalo central a posteriori. Las preferencias de una u otra región depende muchas veces de cada autor. Robert (1994), presenta una visión general de éstas y otras posibles regiones y revisa criterios de selección relacionados con la utilidad de tales regiones a través de funciones de pérdida. Tanner (1996) destaca la utilización de las regiones *HPD* y propone

métodos computacionales para su cálculo basados en la obtención de percentiles sobre los valores de $\mathbf{p}_n(\mathbf{q})$.

4.1.1 Regiones HPD.

Sea $g_n(\mathbf{q})$ proporcional a la densidad marginal a posteriori para \mathbf{q} después del estadio n , $\mathbf{p}_n(\mathbf{q}) = \mathbf{p}_n(\mathbf{q} | X_n, Y_n, \mathbf{p}_0)$. Sean además g_n^{max} y g_n^{min} los valores máximos y mínimos de $g_n(\mathbf{q})$ sobre el soporte de \mathbf{q} . Dado $t \in [0, 1]$, la región $hpd_n(t)$ queda definida como

$$hpd_n(t) = \{\mathbf{q} : g_n(\mathbf{q}) \geq g_n^{max} - t(g_n^{max} - g_n^{min})\}.$$

En particular, $hpd_n(t=0)$ es una región que sólo contiene a la moda de $\mathbf{p}_n(\mathbf{q})$, \mathbf{q}_n^{pmo} , y $hpd_n(t=1)$ coincidirá con el soporte $\mathbf{p}_n(\mathbf{q})$. La región $hpd_n(t)$ es un intervalo o un conjunto de intervalos disjuntos dependiendo del tipo de modalidad que presente dicha distribución y de t , y el ancho de dicha región será monótonamente creciente con t . Cuando la distribución marginal a posteriori para \mathbf{q} es unimodal, el $hpd_n(t)$ será un intervalo para cualquier t . Tanner (1996) presenta herramientas útiles para el cálculo del contenido y de la frontera de dicha región.

Dado que \mathbf{q}_n^{pmo} pertenece a $hpd_n(t)$ para cualquier t , éste divide la región en dos partes. Sean l_s y l_i las longitudes de la parte más ancha y más estrecha, respectivamente, con la convención de que cuando $hpd_n(t)$ es un conjunto de intervalos disjuntos las longitudes se evalúan sólo considerando los puntos dentro de $hpd_n(t)$. Las Figuras 4.1 y 4.2 ilustran todas estas definiciones. En la Figura 4.1 se muestra el caso de una distribución unimodal y en la Figura 4.2 el de una distribución bimodal, en la que se aprecia que $hpd_n(t)$ puede estar formado por intervalos disjuntos. Los diseños HPD, que serán definidos en la Sección 4.2, experimentarán sobre los $hpd_n(t)$ con lo que estas dos figuras también ayudarán a entenderlos.

4.1.2 Intervalos centrales a posteriori.

Sean $q_n^{a/2}$ y $q_n^{1-a/2}$ los cuantiles $a/2$ y $(1-a/2)$ de la distribución a posteriori para el máximo de $f(x; \mathbf{b})$, \mathbf{k} . Un intervalo central a posteriori para \mathbf{k} se define como $cp_n(\mathbf{a}) = [q_n^{a/2}, q_n^{1-a/2}]$. Notar que $cp_n(\mathbf{a}=1)$ sólo contendrá a la mediana de la distribución $p_n(\mathbf{k})$, \mathbf{k}_n^{pme} , y que $cp_n(\mathbf{a}=0)$ es el soporte de la densidad a posteriori de \mathbf{k} , $p_n(\mathbf{k})$. El ancho de $cp_n(\mathbf{a})$ es monótonamente decreciente en \mathbf{a} y dado que \mathbf{k}_n^{pme} pertenece a $cp_n(\mathbf{a})$ para cualquier \mathbf{a} , esta mediana dividirá el intervalo en dos subintervalos.

Es importante recalcar que así como las regiones *HPD* podrían ser conjuntos de intervalos disjuntos, los intervalos centrales a posteriori no lo pueden ser y así como los $hpd_n(\mathbf{t})$ contienen la moda los $cp_n(\mathbf{a})$ contienen la mediana. Una manera de construir $cp_n(\mathbf{a})$ es a través de la obtención de una muestra de la marginal a posteriori de \mathbf{q} ó \mathbf{k} y utilizar un algoritmo de ordenación o de búsqueda como el propuesto por Knuth (1973) para encontrar los cuantiles.

4.2 Diseños para control multiperiodo basados en regiones *HPD*.

Los diseños *HPD* estarán indexados por $(\mathbf{t}, \mathbf{g}) \in [0, 1] \times [-1, 1]$ y se denotarán por $d_{hpd}(\mathbf{t}, \mathbf{g})$. Ellos observan y_{n+1} en un punto $x_{n+1}^{hpd}(\mathbf{t}, \mathbf{g})$ del conjunto $hpd_n(\mathbf{t})$. El índice \mathbf{t} regula el ancho de la región. Cuanto más grande es \mathbf{t} , más lejos de la elección *certainty equivalence* podrá observar el diseño. El índice \mathbf{g} indica la posición relativa de $x_{n+1}^{hpd}(\mathbf{t}, \mathbf{g})$ en el conjunto $hpd_n(\mathbf{t})$. Cuando $\mathbf{g} > 0$, $d_{hpd}(\mathbf{t}, \mathbf{g})$ experimenta en el lado de \mathbf{q}_n^{pmo} donde $hpd_n(\mathbf{t})$ es más ancho, a una distancia de x_{n+1}^{ce} igual a \mathbf{g} veces la longitud de esa parte, l_s . Cuando

$\mathbf{g} < 0$ se experimenta en el lado más estrecho a una distancia de x_{n+1}^{ce} igual a \mathbf{g} veces su longitud, l_i . Así para un \mathbf{t} fijado, cuanto más grande es $|\mathbf{g}|$, más lejos de \mathbf{q}_n^{pmo} podrá observar el nuevo $x_{n+1}^{hpd}(\mathbf{t}, \mathbf{g})$. Cuando $|\mathbf{g}|=1$, $x_{n+1}^{hpd}(\mathbf{t}, \mathbf{g})$ será uno de los extremos de $hpd_n(\mathbf{t})$, dependiendo del signo de \mathbf{g} .

Las Figuras 4.1 y 4.2 muestran todas estas definiciones para distribuciones marginales a posteriori para \mathbf{q} unimodal y bimodal, respectivamente. En la Figura 4.2, al ser la distribución bimodal existen valores de \mathbf{t} para los que el $hpd_n(\mathbf{t})$ está formado por dos intervalos, y l_s se calcula como la suma de la longitud de los dos segmentos de la región $hpd_n(\mathbf{t})$ a la derecha de \mathbf{q}_n^{pmo} .

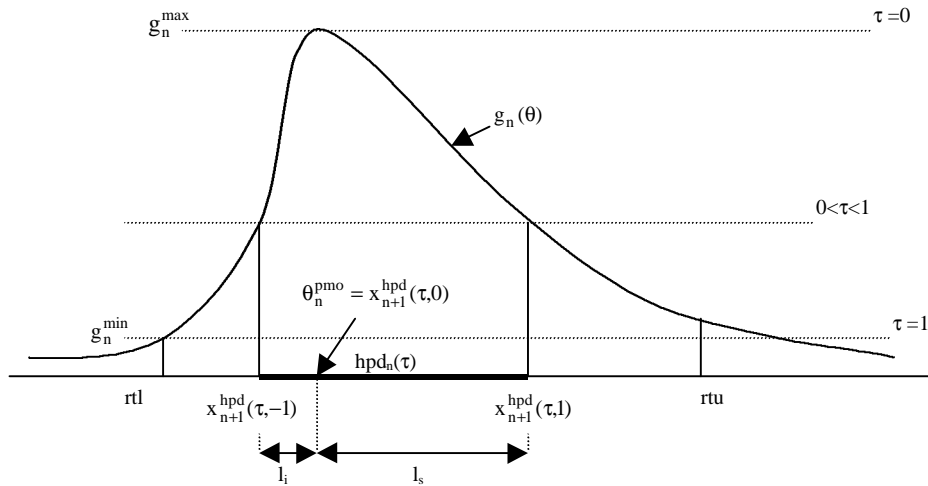


Figura 4.1 Ejemplo de una región $hpd_n(\mathbf{t})$ y de los diferentes ingredientes que intervienen en la definición del diseño $d_{hpd}(\mathbf{t}, \mathbf{g})$ cuando la marginal a posteriori para \mathbf{q} es unimodal y el soporte es $[rtl, rtu]$.

Notar que cuando $\mathbf{t}=0$ ó $\mathbf{g}=0$, $d_{hpd}(\mathbf{t}, \mathbf{g})$ experimenta en $x_{n+1}^{hpd} = x_{n+1}^{ce} = \mathbf{q}_n^{pmo}$. Cuando n aumenta, $hpd_n(\mathbf{t})$ tenderá a ser más estrecho y $x_{n+1}^{hpd}(\mathbf{t}, \mathbf{g})$ se aproxima a \mathbf{q}_n^{pmo} de acuerdo con la necesidad decreciente de perturbar *certainty equivalence* para aprender sobre \mathbf{q} a medida que n se acerca a N .

$T + \mathbf{b}_1(x - \mathbf{q})$ con $\mathbf{q} = (T - \mathbf{b}_0)/\mathbf{b}_1 \sim U(-R, R)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{\mathbf{b}_1} = 1., \mathbf{s}_{\mathbf{b}_1}^2)$ y donde \mathbf{b}_1 es independiente de \mathbf{q} . También se supondrá sin pérdida de generalidad que $T = 0$. La distribución marginal a posteriori para \mathbf{q} como la ecuación que verifican las modas de dicha distribución aparecen en el Apéndice 4.A.

Para estimar $r_2(\mathbf{t}, \mathbf{g})$ se simularán 500.000 realizaciones de $(\mathbf{b}_1, \mathbf{q})$, y para cada $(\mathbf{b}_1, \mathbf{q})$ se ejecutará el $d_{hpd}(\mathbf{t}, \mathbf{g})$ sobre las y_i 's simuladas de $(y_i | x_i^{hpd}, \mathbf{b}_1, \mathbf{q})$ promediando sus pérdidas observadas. Como la desviación estándar estimada para $\hat{r}_2(\mathbf{t}, \mathbf{g})$ es del orden de .000893, el $(\mathbf{t}^*, \mathbf{g}^*)$ minimizando $r_2(\mathbf{t}, \mathbf{g})$ puede ser fácilmente localizado con el algoritmo dado por Nelder and Mead (1965).

Para evaluar la mejora relativa de $d_{hpd}(\mathbf{t}^*, \mathbf{g}^*)$ con respecto a cualquier otro diseño *HPD*, se define $rr(\mathbf{t}, \mathbf{g})$ como la diferencia relativa entre la pérdida esperada bajo el diseño $d_{hpd}(\mathbf{t}^*, \mathbf{g}^*)$ y la esperada bajo $d_{hpd}(\mathbf{t}, \mathbf{g})$,

$$rr(\mathbf{t}, \mathbf{g}) = \frac{r_2(\mathbf{t}, \mathbf{g}) - r_2(\mathbf{t}^*, \mathbf{g}^*)}{r_2(\mathbf{t}^*, \mathbf{g}^*)}.$$

En particular, $rr(0, \mathbf{g})$ es la mejora relativa de $d_{hpd}(\mathbf{t}^*, \mathbf{g}^*)$ sobre el diseño *certainty equivalence* que experimenta en $x_{n+1}^{hpd}(0, \mathbf{g}) = x_{n+1}^{ce} = \mathbf{q}_n^{pmo}$.

La Figura 4.3 explora como \mathbf{t}^* , \mathbf{g}^* y $rr(0, \mathbf{g})$ dependen de $\mathbf{s}_{\mathbf{b}_1}$, de \mathbf{s} , del horizonte N y del tamaño de la región de soporte para \mathbf{q} , R . Si $N = 8$ y $R = .5$, \mathbf{t}^* y \mathbf{g}^* aumentan cuando lo hace \mathbf{s} , mientras son bastantes insensibles a las variaciones en $\mathbf{s}_{\mathbf{b}_1}$. Para la mayoría de las combinaciones de $(\mathbf{s}_{\mathbf{b}_1}, \mathbf{s})$ probadas, \mathbf{g}^* fue positivo y esto significa que el mejor diseño *HPD* experimenta en la parte más ancha de $hpd_n(\mathbf{t})$.

Por otra parte, el comportamiento en función de N y R de \mathbf{t}^* y de \mathbf{g}^* son similares; cuando N aumenta, tanto \mathbf{t}^* como \mathbf{g}^* disminuyen con lo que para N grande el mejor diseño *HPD* estará cerca del diseño *certainty equivalence*. Cuando R disminuye, \mathbf{t}^* y \mathbf{g}^* primero

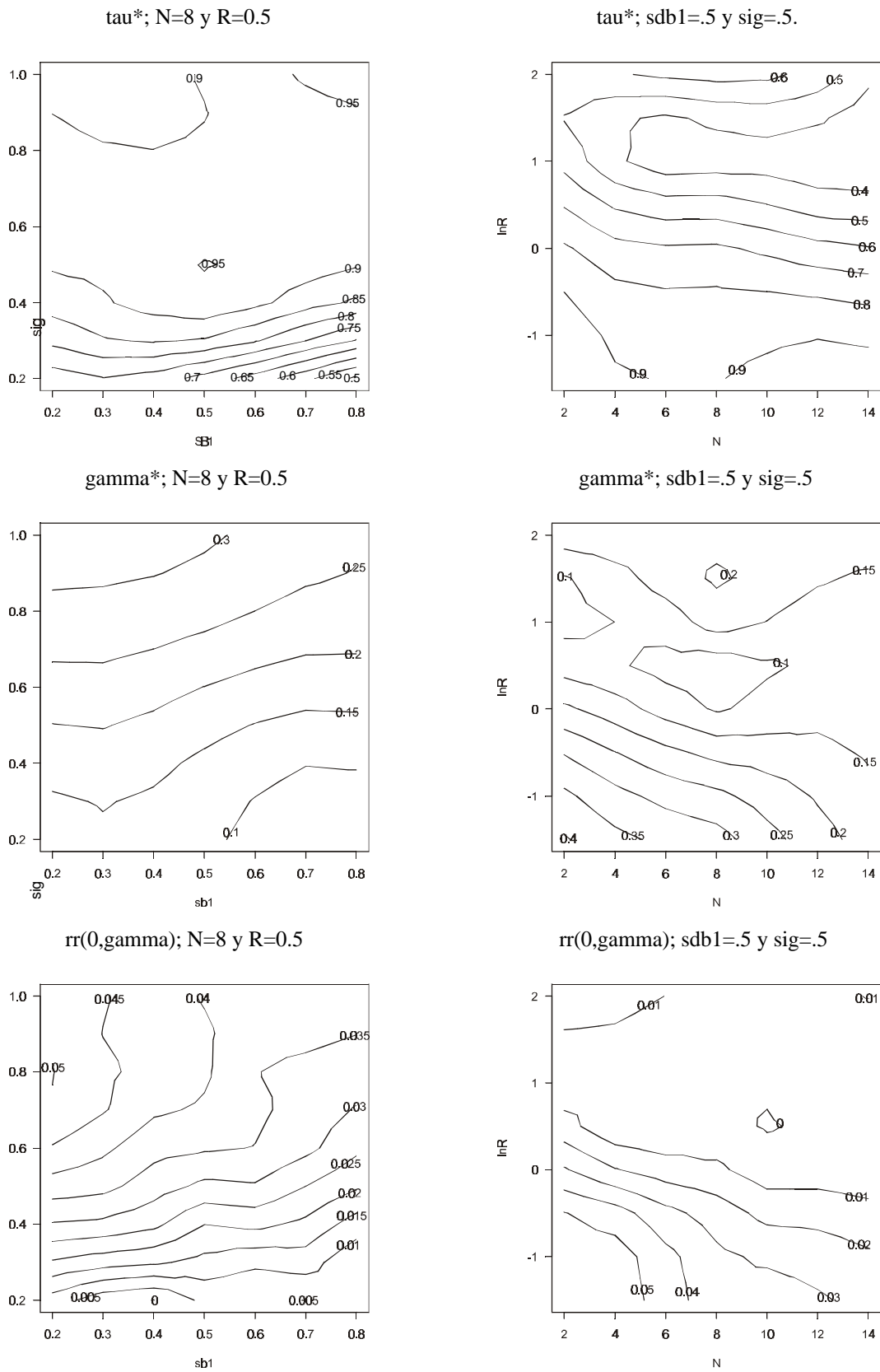


Figura 4.3. Gráficos de contorno de \mathbf{t}^* , \mathbf{g}^* y de la mejora relativa de $d_{hpd}(\mathbf{t}^*, \mathbf{g}^*)$ sobre c.e., $rr(\mathbf{t} = \mathbf{0}, \mathbf{g})$, cuando $E[y | x, \mathbf{b}] = T + \mathbf{b}_1(x - \mathbf{q})$ con $\mathbf{q} \sim U(-R, R)$, $\mathbf{b}_1 \sim N(1, \mathbf{s}_{b_1}^2)$ y \mathbf{b}_1 independiente de \mathbf{q} .

disminuyen para luego aumentar. Este comportamiento se debe al efecto frontera. La probabilidad de que la moda de la distribución a posteriori esté en un extremo de la región crece cuando R disminuye, con lo que al aumentar \mathbf{t}^* y \mathbf{g}^* se fomenta la observación en puntos interiores.

La Figura 4.3 también muestra que $rr(0, \mathbf{g})$ aumenta cuando aumenta \mathbf{s} y disminuye \mathbf{s}_{b_1} , siendo más sensible a las variaciones en \mathbf{s} que en \mathbf{s}_{b_1} . Esto significa que cuanto más se sabe sobre \mathbf{b}_1 y mayor es la variabilidad de y sobre su media, mayor es la mejora relativa de $d_{hpd}(\mathbf{t}^*, \mathbf{g}^*)$ sobre *certainty equivalence*. Por otra parte, $rr(0, \mathbf{g})$ disminuye cuando aumenta N , como era de esperar del hecho que Hu (1997) y Chen y Hu (1998) demuestran que *certainty equivalence* Bayes es asintóticamente óptimo y, por tanto, es eficiente para N muy grande. En resumen, se ha encontrado que el comportamiento de \mathbf{t}^* , \mathbf{g}^* y $rr(0, \mathbf{g})$ con respecto a \mathbf{s}_{b_1} , \mathbf{s} , N y R es similar para todas las prioris, horizontes y tamaños del soporte de \mathbf{q} que se han probado.

4.4 Diseños para el *RSB* basados en intervalos centrales a posteriori.

Los diseños *certainty equivalence* para el *RSB* observan y_{n+1} en algún valor estimado para el máximo en x de $f(x; \mathbf{b})$, $x_{n+1}^{ce} = \hat{\mathbf{k}}_n$. Ginebra y Clayton (1995) perturban *certainty equivalence* experimentando sobre el x_{n+1} maximizando una cota superior de la superficie prevista, $\hat{f}_n(x; \mathbf{b})$. En esta sección se describe la posibilidad de experimentar en intervalos centrales a posteriori para el máximo de $f(x; \mathbf{b})$, \mathbf{k} , perturbando los diseños *certainty*

equivalence que observan y_{n+1} en la mediana de la distribución a posteriori para \mathbf{k} , $x_{n+1}^{ce} = \mathbf{k}_n^{pme}$.

Los diseños basados en los intervalos centrales a posteriori, que a partir de ahora se denominarán diseños *CP*, están indexados por los parámetros $(\mathbf{a}, \mathbf{g}) \in [0, 1] \times [-1, 1]$ y serán denotados por $d_{cp}(\mathbf{a}, \mathbf{g})$. Ellos observan y_{n+1} en un punto $x_{n+1}^{cp}(\mathbf{a}, \mathbf{g})$ sobre el intervalo $cp_n(\mathbf{a})$. Cuanto menor es \mathbf{a} , más ancho es $cp_n(\mathbf{a})$ y, por lo tanto, más lejos de la elección *certainty equivalence* podría observar este diseño. El punto \mathbf{k}_n^{pme} divide el intervalo $cp_n(\mathbf{a})$ en dos subintervalos. El índice \mathbf{g} indica la posición relativa de $x_{n+1}^{cp}(\mathbf{a}, \mathbf{g})$ sobre ese intervalo. Cuando $\mathbf{g} > 0$, $d_{cp}(\mathbf{a}, \mathbf{g})$ experimenta en el subintervalo más ancho, a una distancia de $x_{n+1}^{ce} = \mathbf{k}_n^{pme}$ igual a \mathbf{g} veces la longitud de ese subintervalo. Cuando $\mathbf{g} < 0$, $d_{cp}(\mathbf{a}, \mathbf{g})$ experimentará en el subintervalo más estrecho, a una distancia de $x_{n+1}^{ce} = \mathbf{k}_n^{pme}$ igual a $|\mathbf{g}|$ veces la longitud de ese subintervalo. Así, para un valor de \mathbf{a} fijo, cuanto mayor es $|\mathbf{g}|$ más lejos de \mathbf{k}_n^{pme} estará $x_{n+1}^{cp}(\mathbf{a}, \mathbf{g})$. Cuando $|\mathbf{g}| = 1$, $x_{n+1}^{cp}(\mathbf{a}, \mathbf{g})$ es $q_n^{a/2}$ ó $q_n^{1-a/2}$ dependiendo de cual de estos cuantiles está más cerca de \mathbf{k}_n^{pme} .

Cuando $\mathbf{a} = 1$ ó $\mathbf{g} = 0$, $d_{cp}(\mathbf{a}, \mathbf{g})$ experimenta en $x_{n+1}^{cp} = x_{n+1}^{ce} = \mathbf{k}_n^{pme}$. Cuando n aumenta, $x_{n+1}^{cp}(\mathbf{a}, \mathbf{g})$ tenderá a estar más cerca de \mathbf{k}_n^{pme} , pues la distribución a posteriori estará más concentrada entorno a \mathbf{k}_n^{pme} , incorporando así la decreciente necesidad de aprender sobre \mathbf{k} cuando el valor de n está próximo a N .

Para encontrar el diseño $d_{cp}(\mathbf{a}^*, \mathbf{g}^*)$ que maximiza $w(\mathbf{a}, \mathbf{g}) = w(d_{cp}(\mathbf{a}, \mathbf{g})) = E[\sum_{n=1}^N y_n(x_n^{cp})]$ se procederá de forma análoga a la descrita en la Sección 4.2 para los diseños *HPD*. Para cada (\mathbf{a}, \mathbf{g}) se simula repetidamente el uso del correspondiente diseño sobre datos del modelo, estimando su utilidad esperada, $w(\mathbf{a}, \mathbf{g})$. Buscando entre los distintos $d_{cp}(\mathbf{a}, \mathbf{g})$ se obtiene el $d_{cp}(\mathbf{a}^*, \mathbf{g}^*)$ que maximiza $w(\mathbf{a}, \mathbf{g})$, que siempre mejorará al diseño *certainty equivalence* que experimenta en $x_{n+1}^{ce} = \mathbf{k}_n^{pme}$. Cuando \mathbf{a}^* está cerca de 1 ó \mathbf{g}^* está cerca de 0, ese diseño *certainty equivalence* será el mejor de la familia.

4.5 Aplicación de los diseños “CP” al modelo lineal.

Para ilustrar el uso de estos diseños se realiza un estudio de sensibilidad para un modelo lineal normal con priori conjugada. Las propiedades distribucionales para esta situación se describen en la Sección 1 del Apéndice I. En este ejemplo, $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \mathbf{b}_1x + \mathbf{b}_2x^2$ y $\mathbf{k} = -\mathbf{b}_1/(2\mathbf{b}_2)$, con las \mathbf{b}_i 's independientes entre sí y $\mathbf{b}_0 \sim N(1., \mathbf{s}_{b_0}^2 = 1.)$, $\mathbf{b}_1 \sim N(0., \mathbf{s}_{b_1}^2)$ y $\mathbf{b}_2 \sim N(-1., \mathbf{s}_{b_2}^2 = .2)$. La región experimental es $C = [-R, R]$ y cuando el punto en el que se debería observar según los valores de \mathbf{a} y \mathbf{g} cae fuera de C , se experimentará en el extremo de C más cercano a ese punto.

Para estimar $w(\mathbf{a}, \mathbf{g})$ se simula 150.000 realizaciones de $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2)$ a partir de la distribución a priori para \mathbf{b} , y para cada \mathbf{b} simulada se ejecuta el diseño $d_{cp}(\mathbf{a}, \mathbf{g})$ sobre y_i 's simuladas a partir de la distribución de $(y_i | x_i^{cp}, \mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2)$, promediando las utilidades observadas. Como la desviación estándar estimada para $\hat{w}(\mathbf{a}, \mathbf{g})$ es del orden de 0,009, el $(\mathbf{a}^*, \mathbf{g}^*)$ que maximiza $w(\mathbf{a}, \mathbf{g})$ puede ser localizado fácilmente.

El gasto computacional resulta ser bastante elevado y ejecutarlo para horizontes N superiores a 16 implicó reducir notablemente el número de simulaciones de \mathbf{b} y con ello la pérdida de precisión en los valores estimados de la utilidad esperada. En cada estadio es preciso obtener una muestra de la distribución a posteriori de \mathbf{k} y calcular los cuantiles $q_n^{a/2}$ y $q_n^{1-a/2}$. Para hallar estos cuantiles reduciendo el tiempo de computación se empleó un método de búsqueda, que no requiere la ordenación completa de la muestra, dado por Knuth (1973) e implementado en el libro de Press et al (1992).

Para evaluar la mejora relativa de $d_{cp}(\mathbf{a}^*, \mathbf{g}^*)$ con respecto a $d_{cp}(\mathbf{a}, \mathbf{g})$, se define $rw(\mathbf{a}, \mathbf{g})$ como la diferencia relativa entre la utilidad esperada bajo $d_{cp}(\mathbf{a}^*, \mathbf{g}^*)$ y la esperada bajo $d_{cp}(\mathbf{a}, \mathbf{g})$,

$$rw(\mathbf{a}, \mathbf{g}) = \frac{w(\mathbf{a}^*, \mathbf{g}^*) - w(\mathbf{a}, \mathbf{g})}{w(\mathbf{a}^*, \mathbf{g}^*)}.$$

En particular, $rw(\mathbf{a}=1, \mathbf{g})$ es la mejora relativa de $d_{cp}(\mathbf{a}^*, \mathbf{g}^*)$ con respecto al diseño *certainty equivalence* que experimenta en $x_{n+1}^{cp}(1, \mathbf{g}) = x_{n+1}^{ce} = \mathbf{k}_n^{pme}$.

La Figura 4.4 explora como \mathbf{a}^* , \mathbf{g}^* y $rw(\mathbf{a}=1, \mathbf{g})$ dependen de \mathbf{s}_{b_1} , de \mathbf{s} , del horizonte N y del tamaño de la región experimental R . Cuando $N=10$ y $R=1.$, \mathbf{a}^* y \mathbf{g}^* disminuyen cuando aumenta \mathbf{s}_{b_1} , mientras son algo menos sensibles a las variaciones de \mathbf{s} . Para la mayoría de las combinaciones de $(\mathbf{s}_{b_1}, \mathbf{s})$ probadas, \mathbf{g}^* fue negativo y el mejor diseño *CP* experimenta en el subintervalo más estrecho del $cp_n(\mathbf{a})$. Por otra parte, \mathbf{a}^* y \mathbf{g}^* son relativamente poco sensibles a las variaciones de N . Cuando aumenta R , \mathbf{a}^* aumenta mientras que \mathbf{g}^* primero aumenta y luego disminuye.

La Figura 4.4 también muestra que $rw(1, \mathbf{g})$ aumenta cuando aumenta \mathbf{s}_{b_1} y disminuye \mathbf{s} . Cuanto menos se sabe sobre \mathbf{b}_1 y menor es la variabilidad de \mathbf{y} mayor será la mejora sobre *certainty equivalence*. Por otra parte, para el rango de N probadas $rw(1, \mathbf{g})$ aumenta cuando aumenta N , indicando que la mejora del diseño *CP* respecto a nuestro *certainty equivalence* aumenta al aumentar el horizonte. Este comportamiento es el contrario al que se observa al implementar *Fieller* y *HPD* para problemas de control.

4.6 Extensiones y comentarios.

Las mejoras alcanzadas con $d_{hpd}(\mathbf{t}^*, \mathbf{g}^*)$ sobre el diseño *certainty equivalence* que experimenta sobre $x_{n+1}^{ce} = \mathbf{q}_n^{pmo}$ en los ejemplos presentados en la Sección 4.3 son de orden algo inferior a las obtenidas con los diseños expuestos en el Capítulo 3. Se ha comprobado que para estos ejemplos este diseño *certainty equivalence* ya es muy bueno. Las mejoras para el caso de los diseños $d_{cp}(\mathbf{a}^*, \mathbf{g}^*)$ si son bastante más substanciales. Una forma de mejorar $d_{hpd}(\mathbf{t}^*, \mathbf{g}^*)$ y

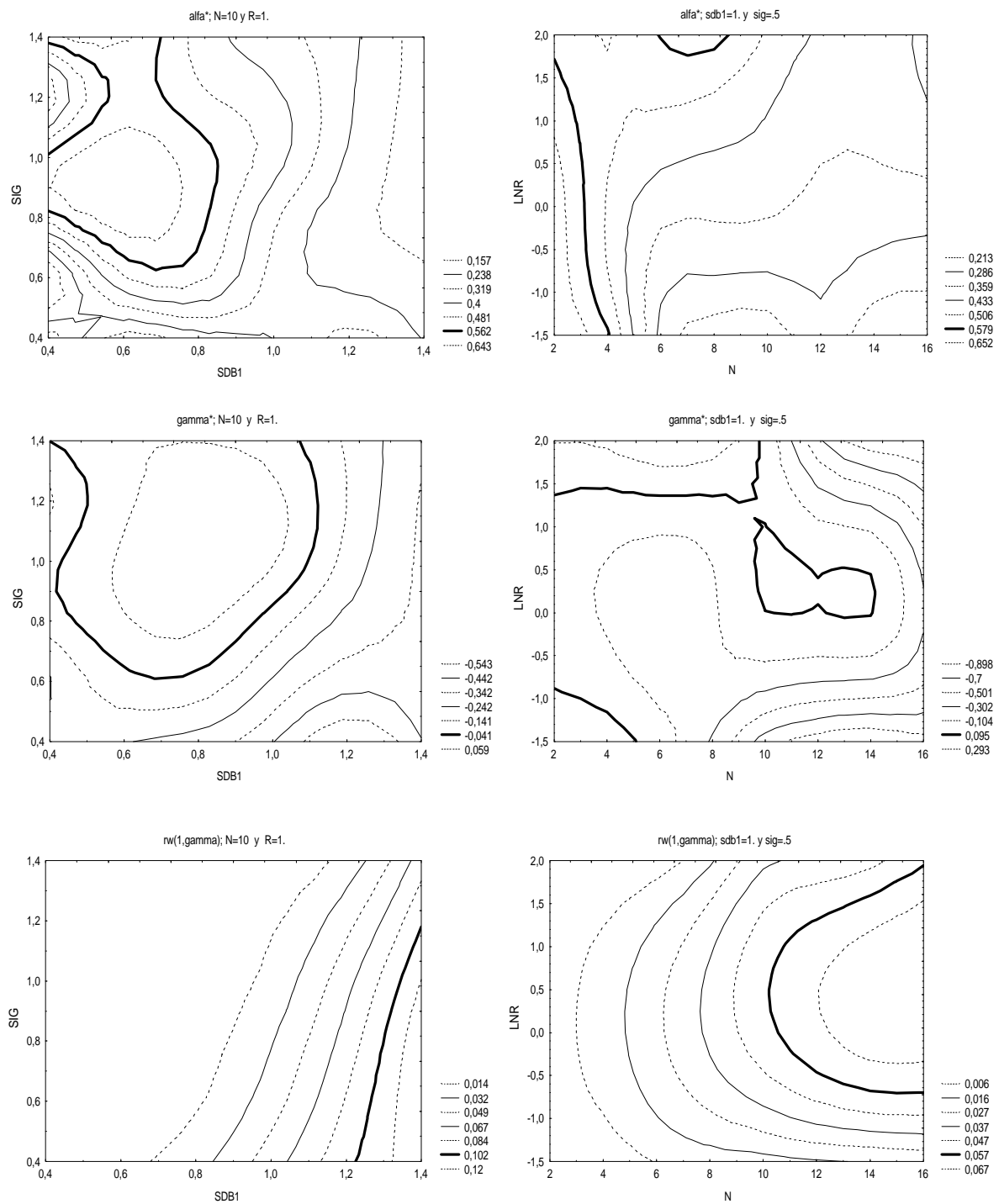


Figura 4.4. Gráficos de contorno para \mathbf{a}^* , \mathbf{g}^* y de la mejora relativa de $d_{cp}(\mathbf{a}^*, \mathbf{g}^*)$ sobre $d_{cp}(1, \mathbf{g})$, $rw(1, \mathbf{g})$, cuando $E[y | x, \mathbf{b}] = \mathbf{b}_0 + \mathbf{b}_1 x + \mathbf{b}_2 x^2$, con las \mathbf{b}_i 's independientes y $\mathbf{b}_0 \sim N(1., \mathbf{s}_{\mathbf{b}_0}^2 = 1.)$, $\mathbf{b}_1 \sim N(0., \mathbf{s}_{\mathbf{b}_1}^2)$, $\mathbf{b}_2 \sim N(-1., \mathbf{s}_{\mathbf{b}_2}^2 = .2)$ y con $C = [-R, R]$.

$d_{cp}(\mathbf{a}^*, \mathbf{g}^*)$ sería reestimar \mathbf{t}^* ó \mathbf{a}^* y \mathbf{g}^* en algunos m estadios intermedios, basándose en las distribuciones a posteriori para \mathbf{q} y \mathbf{k} , respectivamente.

La principal dificultad cuando se implementan los diseños propuestos es la computación repetida de dichas regiones creíbles para estimar $(\mathbf{t}^*, \mathbf{g}^*)$ ó $(\mathbf{a}^*, \mathbf{g}^*)$. Usando diseños tipo *HPD* se requiere de la computación repetida del perfil de la marginal a posteriori de \mathbf{q} ó \mathbf{k} . Cuando no se puede disponer de ese perfil, se puede implementar los diseños basados en intervalos centrales a posteriori, donde la dificultad está en la obtención de una muestra relativamente amplia de la marginal a posteriori y en la evaluación repetida de los cuantiles a posteriori de \mathbf{q} ó \mathbf{k} , a través de algoritmos de búsqueda. Cuando ambas posibilidades no son factibles, podrían utilizarse los diseños propuestos en el Capítulo 3 y los que se propondrán en los capítulos siguientes, a la vez que los dados por Ginebra y Clayton (1995). Una buena característica de los diseños propuestos en este capítulo con respecto a los de los capítulos anteriores es que sus índices a parte de estar acotados de forma natural están relacionados con conceptos básicos de la Estadística.

La misma idea usada para definir los diseños propuestos en este capítulo podría ser generalizada a otros problemas secuenciales con una variable de entrada. Su extensión a problemas con más de una variable de entrada no sería nada sencillo, entre otras cosas, porque cuando $x \in R^k$, la solución de $f(x; \mathbf{b}) = T$ será un hiperplano y no un número real, y \mathbf{k} un punto en R^k .

Por último notar que en este capítulo (y en toda la memoria) se seleccionan los mejores diseños dentro de la familia usando el mismo modelo y priori que los utilizados para construir la familia de diseños. Esto no es obligatorio hacerlo siempre. En particular, se podría utilizar el modelo lineal para generar la familia de diseños pero simular las repuestas a partir de un modelo más complicado, pero más realista a la hora de calcular $(\mathbf{t}^*, \mathbf{a}^*)$ ó $(\mathbf{a}^*, \mathbf{g}^*)$. Por ejemplo, esto podría ser adecuado para datos procedentes de estudios de supervivencia.

Apéndice 4.A Marginal de q para el ejemplo de la Sección 4.3.

Para implementar los diseños *HPD* para el modelo lineal normal con las distribuciones a priori usadas en la Sección 4.3, donde $y_n | x_n, \mathbf{b} \sim N(\mathbf{b}_0 + \mathbf{b}_1 x_n, \mathbf{s}^2)$, de manera que $E[y | x, \mathbf{b}] = \mathbf{b}_0 + \mathbf{b}_1 x = T + \mathbf{b}_1(x - \mathbf{q})$, con $\mathbf{q} = (T - \mathbf{b}_0) / \mathbf{b}_1$, se utiliza la siguiente proposición:

Proposición 4.1 Sean $y'_n = y_n - T$ con $y'_n | x_n, \mathbf{b} \sim N(\mathbf{b}_1(x_n - \mathbf{q}), \mathbf{s}^2)$, $\mathbf{q} \sim U(\text{rtl}, \text{rtu})$, $\mathbf{b}_1 \sim N(\mathbf{m}_{\mathbf{b}_1}, \mathbf{s}_{\mathbf{b}_1}^2)$, \mathbf{b}_1 independiente de \mathbf{q} y \mathbf{s} conocida. La densidad marginal a posteriori para \mathbf{q} es

$$\mathbf{p}_n(\mathbf{q}) \propto \frac{I_{[\text{rtl}, \text{rtu}]}(\mathbf{q})}{\sqrt{\mathbf{s}_{\mathbf{b}_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2}} \exp \left\{ \frac{(\mathbf{s}_{\mathbf{b}_1}^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 \mathbf{m}_{\mathbf{b}_1})^2}{2\mathbf{s}^2 \mathbf{s}_{\mathbf{b}_1}^2 (\mathbf{s}_{\mathbf{b}_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2)} \right\}.$$

Su demostración aparece en la Sección 1 del Apéndice II. La siguiente proposición muestra como la marginal a posteriori para \mathbf{q} tiene a lo sumo dos modas.

Proposición 4.2 Sean $y'_n = y_n - T$ con $y'_n | x_n, \mathbf{b} \sim N(\mathbf{b}_1(x_n - \mathbf{q}), \mathbf{s}^2)$, $\mathbf{q} \sim U(\text{rtl}, \text{rtu})$, $\mathbf{b}_1 \sim N(\mathbf{m}_{\mathbf{b}_1}, \mathbf{s}_{\mathbf{b}_1}^2)$, \mathbf{b}_1 independiente de \mathbf{q} y \mathbf{s} conocida. La densidad marginal a posteriori para \mathbf{q} tiene a lo sumo dos modas siendo raíces del polinomio o los extremos de $[\text{rtl}, \text{rtu}]$

$$\begin{aligned} & \mathbf{q}^3 (\mathbf{s}_{\mathbf{b}_1}^4 \mathbf{s}^2 n^2) + \mathbf{q}^2 (-3n \mathbf{s}_{\mathbf{b}_1}^4 \mathbf{s}^2 \sum x_i - n \text{cte}2 \mathbf{s}_{\mathbf{b}_1}^2 \sum y_i + \mathbf{s}_{\mathbf{b}_1}^4 (\sum y_i)^2 \sum x_i) \\ & + \mathbf{q} (n \text{cte}2^2 + n \mathbf{s}_{\mathbf{b}_1}^2 \mathbf{s}^2 \text{cte}1 + 2\mathbf{s}_{\mathbf{b}_1}^4 \mathbf{s}^2 (\sum x_i)^2 - \text{cte}1 \mathbf{s}_{\mathbf{b}_1}^2 (\sum y_i)^2) \\ & - \sum x_i (\text{cte}2^2 + \mathbf{s}_{\mathbf{b}_1}^2 \mathbf{s}^2 \text{cte}1) + \text{cte}1 \text{cte}2 \sum y_i = 0. \end{aligned}$$

con $\text{cte}1 = \mathbf{s}_{\mathbf{b}_1}^2 \sum x_i^2 + \mathbf{s}^2$ y $\text{cte}2 = \mathbf{s}_{\mathbf{b}_1}^2 \sum x_i y_i + \mathbf{s}^2 \mathbf{m}_{\mathbf{b}_1}$.

Su demostración sigue de igualar a cero $d\mathbf{p}_n(\mathbf{q})/d\mathbf{q}$ y comprobar que se trata de un polinomio de tercer grado. Observar que $\mathbf{p}_n(\mathbf{q}) = I_{[\text{rtl}, \text{rtu}]}(\mathbf{q}) h_n(\mathbf{q})$ donde $h_n(\mathbf{q})$ tiene una asíntota horizontal en el eje de abscisas. Si las tres raíces del polinomio fueran reales, tendrá

dos máximos y un mínimo local entre ellos, faltando sólo por comprobar si dichos máximos pertenecen al intervalo $[rtl, rtu]$. Si ambos máximos pertenecen a $[rtl, rtu]$, serán las dos modas de $\mathbf{p}_n(\mathbf{q})$. Si el mínimo local y uno de los máximos pertenecen a $[rtl, rtu]$, $\mathbf{p}_n(\mathbf{q})$ también será bimodal, ya que una moda será dicho máximo y la otra uno de los extremos del intervalo $[rtl, rtu]$. Si el mínimo local es el único punto crítico que pertenece $[rtl, rtu]$, $\mathbf{p}_n(\mathbf{q})$ también será bimodal y sus modas coinciden con los extremos del intervalo. En el resto de los casos $\mathbf{p}_n(\mathbf{q})$ será unimodal.

En los ejemplos implementados en la Sección 4.3 se llegan a producir distribuciones bimodales, es decir, situaciones para las que el polinomio anterior tiene tres soluciones reales dentro del soporte de \mathbf{q} , en los primeros estadios con n pequeño. Cuando esto ocurre, una de las modas tiende a ser de varios órdenes de magnitud mayores que la otra, por lo que a efectos prácticos el $hpd_n(\mathbf{t})$ está constituido por un único intervalo para casi todos valores de $\mathbf{t} \in [0, 1]$.

Notar que en el estadio cero, la distribución sobre \mathbf{q} es uniforme, lo que hace que el $hpd_0(\mathbf{t})$ coincida con el soporte de \mathbf{q} para cualquier \mathbf{t} y, por tanto, $x_1^{hpd}(\mathbf{t}, \mathbf{g}) = \mathbf{q}_0^{pmo} = (rtl + rtu)/2$. Si el soporte de \mathbf{q} está centrado respecto a cero, esto hace que $x_1^{hpd}(\mathbf{t}, \mathbf{g}) = 0$ y el polinomio de tercer grado a resolver para la obtención de las modas en el estadio uno quede como

$$\mathbf{q}^3 \mathbf{s}_{b_1}^4 + \mathbf{q}^2 \mathbf{m}_{b_1} \mathbf{s}_{b_1}^2 y_1 + \mathbf{q} (\mathbf{s}^2 (\mathbf{m}_{b_1}^2 + \mathbf{s}_{b_1}^2) - \mathbf{s}_{b_1}^2 y_1^2) + \mathbf{s}^2 \mathbf{m}_{b_1} y_1 = 0,$$

con y_1 el valor observado en $x_1^{hpd}(\mathbf{t}, \mathbf{g}) = 0$.

CAPITULO 5.

DISEÑOS QUE ADAPTAN LAS SECUENCIAS DE APROXIMACION ESTOCÁSTICA (DISEÑOS *SAN*).

Como ya se comentó en los Capítulos 3 y 4, los diseños *certainty equivalence* no toman en consideración la incertidumbre en la estimación de la superficie y por ello en dichos capítulos se propuso la familia de diseños *Fieller* y familia de diseños *HPD* que perturban tal diseño con la idea de aprender activamente sobre la superficie y conseguir así un mejor control futuro. Los diseños *Fieller* obligan en cada estadio a estimar la desviación típica de los valores previstos, $\hat{s.d.}(\hat{y}_n(x))$, y al cálculo de las raíces de $\hat{y}_n(x) \pm K \hat{s.d.}(\hat{y}_n(x)) = T$. Los diseños *HPD* en cambio necesitan del cálculo de las regiones de máxima densidad a posteriori en cada uno de los estadios. Estos cálculos no siempre son sencillos, por lo que en este capítulo se introduce otra familia de diseños, que llamaremos diseños *san*, con un comportamiento similar a *Fieller* pero que no obligan a calcular $\hat{s.d.}(\hat{y}_n(x))$ y que incluyen como miembros de la familia tanto al diseño *certainty equivalence* secuencial como al no secuencial.

La construcción de estos diseños *san* se basa en un procedimiento recursivo en el que sólo es preciso obtener un estimador de la derivada de la función f en la raíz, $f'(\mathbf{q})$, y la estimación de la superficie en un punto, $\hat{y}(x)$. Esta familia de diseños es una modificación de las secuencias de aproximación estocástica introducidas por Robbins y Monro (1951) y cuya

definición se recoge en la Sección 5.1. En la Sección 5.2 se definirá e interpretará dicha familia de diseños indicando además la forma de encontrar el mejor diseño dentro de la familia. En la Sección 5.3 se ejemplifica el comportamiento de los diseños *san* sobre el modelo lineal y en la Sección 5.4 se realiza un análisis de sensibilidad en el que se estudia la mejora con respecto a *certainty equivalence* del mejor diseño de la familia. En la Sección 5.5 se extienden y modifican tales diseños, indicando la posibilidad de generalización a problemas con más de una variable de control. En el apéndice final se recogen las demostraciones de los resultados expuestos a lo largo del capítulo.

5.1 Secuencias de aproximación estocásticas.

Robbins y Monro (1951) proponen estimar la raíz \mathbf{q} , a través de una secuencia recursiva de variables aleatoria, $x_{n+1} = x_n - a_n y_n(x_n)$, donde $\{a_n\}$ es una sucesión tal que $\sum_{n=1}^{\infty} a_n^2 < \infty$ y $\sum_{n=1}^{\infty} a_n = \infty$. Bajo ciertas condiciones de regularidad sobre f , \mathbf{e} y a_n , que son fáciles de satisfacer, demostraron que x_n converge casi seguro a \mathbf{q} y que la distribución límite es normal. En particular, si

$$n^{\mathbf{a}} a_n \xrightarrow{a.s.} A > 0$$

para $\mathbf{a} \in (1/2, 1]$, entonces

$$n^{\mathbf{a}/2} (x_n - \mathbf{q}) \xrightarrow{d} N(0, \mathbf{s}^2(A, \mathbf{a})).$$

Eligiendo $a_n = A/n$ se consigue la mejor razón de convergencia para x_n y, en ese caso, $A = 1/f'(\mathbf{q})$ minimizará la varianza asintótica de x_n . Para un resumen del estado del arte de este área ver Ruppert (1991).

Lai y Robbins (1979, 82) sugieren estimar $f'(\mathbf{q})$ a partir del estimador mínimo cuadrático de la pendiente obtenida al ajustar el modelo $y = \mathbf{b}_0 + \mathbf{b}_1 x + \mathbf{e}$ a los puntos (x_i, y_i) ,

$i = 1, \dots, n$. Ginebra (1996) adapta estas secuencias a los problemas de muestras pequeñas, definiendo el diseño d_{sa} como el que observa y_{n+1} en el punto $x_{n+1}^{sa}(x_1, x_2, \mathbf{a}, a_0) \in C$ que está más cerca de

$$x_{n+1} = x_n^{sa}(x_1, x_2, \mathbf{a}, a_0) - \frac{a_0}{\hat{\mathbf{b}}_{1n}^{ls} n^a} y_n(x_n^{sa}),$$

con x_1 y x_2 dos puntos seleccionados inicialmente. Minimizando $r_i(d_{sa})$ de una forma análoga al caso de los diseños *Fieller* y *HPD* se obtiene el mejor diseño dentro de esa familia. Un inconveniente de estas secuencias es que un resultado de Wu (1986) demuestra que la regla *certainty equivalence* secuencial para el modelo lineal simple con $\hat{\mathbf{b}}_{in} = \hat{\mathbf{b}}_{in}^{ls}$, no es un miembro de esta familia definida en Ginebra (1996). En estas circunstancias, Wu reescribe su regla *certainty equivalence* para $T = 0$ como

$$x_{n+1}^{ce} = \hat{\mathbf{q}}_n = -\frac{\hat{\mathbf{b}}_{0n}}{\hat{\mathbf{b}}_{1n}} = x_n - \frac{1}{n\hat{\mathbf{b}}_{1n}} \left[1 + \frac{(n-1)^2 (\bar{y}_{n-1} / \hat{\mathbf{b}}_{1,n-1})^2}{n \sum_{i=1}^n (x_i - \bar{x}_n)^2} \right] y_n.$$

Cuando el segundo sumando del corchete es despreciable, *certainty equivalence* será un miembro de la familia dada por Ginebra, pero para tamaños de muestras pequeños o moderados, este término no se puede despreciar. Ya Wu indicaba que dicho término debería ser menos dependiente de la información más antigua, cosa que se conseguiría por ejemplo, reemplazando \bar{y}_{n-1} , $\hat{\mathbf{b}}_{1,n-1}$ y $\sum_{i=1}^n (x_i - \bar{x}_n)^2$ por versiones ponderadas que dieran más peso a las observaciones más recientes.

5.2 Diseños basados en la modificación de las aproximaciones estocásticas.

5.2.1 Definición de diseños *san*.

Para la construcción de la familia de diseños *san* adaptaremos las secuencias de aproximación estocásticas dadas por Robbins y Monro (1951), incorporando información a priori en la estimación de $f'(\mathbf{q}; \mathbf{b})$, reemplazando $y_n(x_n)$ por un estimador de la superficie, $\hat{y}_n(x) = \hat{f}_n(x; \mathbf{b})$, y considerando el rango total de diseños alcanzables con $a_n = a_0/n^a$ para $(\mathbf{a}, a_0) \in [-1, 1] \times [0, 1]$. Notar que no se está interesado en asegurar una rápida razón de convergencia de x_n a \mathbf{q} sino en minimizar la pérdida esperada para un número de experimentos N pequeño o moderado. Se definen los diseños $d_{san}(\mathbf{a}, a_0)$ como sigue

Definición 5.1 *Dados $\hat{y}_n(x)$, $\hat{f}'_n(x; \mathbf{b})$ y $(\mathbf{a}, a_0) \in [-1, 1] \times [0, 1]$, el diseño $d_{san}(\mathbf{a}, a_0)$ observa y_{n+1} en el $x_{n+1}^{san}(\mathbf{a}, a_0) \in C$ más cercano a*

$$x_{n+1}(\mathbf{a}, a_0) = x_n^{san}(\mathbf{a}, a_0) - \frac{a_0}{\hat{f}'_n(\mathbf{q}; \mathbf{b}) n^a} (\hat{y}_n(x_n^{san}) - T), \quad (5.1)$$

para $n = 1, 2, \dots, N - 2$. $x_1^{san}(\mathbf{a}, a_0)$ y $x_N^{san}(\mathbf{a}, a_0)$ son los puntos en C más cercanos a x_1 y a x_N soluciones de $\hat{y}_0(x) = T$ e $\hat{y}_{N-1}(x) = T$, respectivamente.

Los estimadores para $f(x; \mathbf{b})$ y $f'(\mathbf{q}; \mathbf{b})$ podrían ser $\hat{y}_n(x) = E_n[y(x)]$ ó $\hat{y}_n(x) = f(x; \hat{\mathbf{b}}_n)$ y $\hat{f}'_n(\mathbf{q}; \mathbf{b}) = f'(\hat{\mathbf{q}}_n; \hat{\mathbf{b}}_n)$, respectivamente, para algunos estimadores de \mathbf{q} y \mathbf{b} , $\hat{\mathbf{q}}_n$ y $\hat{\mathbf{b}}_n$. Para el modelo lineal simple se puede usar $\hat{y}_n(x) = E_n[\mathbf{b}_0] + E_n[\mathbf{b}_1]x$ y $\hat{f}'_n(\mathbf{q}; \mathbf{b}) = E_n[\mathbf{b}_1]$, aunque ésta no será la única elección posible. Cuando se tenga una mejor aproximación de $f(x; \mathbf{b})$ cerca de \mathbf{q} , como es el caso de las curvas de respuesta *cuantal*, ésta debería usarse para estimar $f'(\mathbf{q}; \mathbf{b})$ e $y(x)$.

Notar que tanto $\hat{f}'_n(\mathbf{q}; \mathbf{b})$ como $\hat{y}_n(x)$ acumulan información de todas las observaciones, que es usada para conducir el diseño. Se podrían usar también estimadores que sólo tuvieran en cuenta las observaciones más recientes o estimadores ponderados donde el mayor peso estuviera en esas últimas observaciones.

Es fácil verificar que para el modelo lineal simple, $x_{n+1}^{ce} = x_{n+1}^{san}(\mathbf{a} = 0, a_0 = 1)$ y que $x_{n+1}^{san}(\mathbf{a}, a_0 = 0)$ es el diseño *certainty equivalence* no secuencial que observa todas las y_n 's en el x que resuelve $\hat{y}_0(x) = T$. Así, para el modelo lineal simple, esta familia incluye el correspondiente diseño secuencial y no secuencial *certainty equivalence*. En la siguiente sección se describen detalles de los diseños *san* que acaban de ilustrar la relación entre dichos diseños y los diseños *certainty equivalence*.

5.2.2 Interpretación de diseños *san*.

En la siguiente proposición se relacionan los diseños $d_{san}(\mathbf{a}, a_0)$ y el diseño *certainty equivalence* para el modelo lineal simple. Su demostración se encuentra en el Apéndice 5.A, junto con algunas consideraciones y resultados sobre la interpretación de dichos diseños.

Proposición 5.1 Sea $x_{n+1}^{san}(\mathbf{a}, a_0)$ definido como en (5.1), con $\hat{y}_n(x_n) = f(x_n, \hat{\mathbf{b}}_n)$ y $\hat{f}'_n(\mathbf{q}; \mathbf{b}) = f'(x_n, \hat{\mathbf{b}}_n)$ y sea x_{n+1}^{ce} definido con el $\hat{y}_n(x)$ calculado a partir de los datos (X'_n, Y'_n) elegidos y observados a través de $d_{san}(\mathbf{a}, a_0)$ y $C = R$. Para el modelo lineal simple y $n = 1, 2, \dots, N - 2$,

$$x_{n+1}^{san}(\mathbf{a}, a_0) = \left(1 - \frac{a_0}{n^{\mathbf{a}}}\right) x_n^{san}(\mathbf{a}, a_0) + \frac{a_0}{n^{\mathbf{a}}} x_{n+1}^{ce} = \sum_{i=1}^{n+1} h_n(\mathbf{a}, a_0, i) x_i^{ce}, \quad (5.2)$$

con

$$h_n(\mathbf{a}, a_0, i=1) = \frac{\prod_{j=1}^n (j^{\mathbf{a}} - a_0)}{(n!)^{\mathbf{a}}},$$

$$h_n(\mathbf{a}, a_0, i) = \frac{n^{\mathbf{a}} - a_0}{n^{\mathbf{a}}} h_{n-1}(\mathbf{a}, a_0, i) = \frac{a_0 [(i-2)!]^{\mathbf{a}} \prod_{j=i}^n (j^{\mathbf{a}} - a_0)}{(n!)^{\mathbf{a}}}, \text{ para } i=2, 3, \dots, n,$$

$$h_n(\mathbf{a}, a_0, i=n+1) = \frac{a_0}{n^{\mathbf{a}}}.$$

También

$$\sum_{i=1}^{n+1} h_n(\mathbf{a}, a_0, i) = 1.$$

Como se puede apreciar según esta proposición, $x_{n+1}^{san}(\mathbf{a}, a_0)$ es un promedio ponderado de todas las anteriores elecciones *certainty equivalence* y, en particular, $x_{n+1}^{san}(\mathbf{a}=1, a_0=1) = \sum_{i=1}^n x_{i+1}^{ce}/n$. Bajo las condiciones de la Proposición 5.1, $x_{n+1}^{san}(\mathbf{a}, a_0)$ puede ser escrito como una combinación convexa de $x_n^{san}(\mathbf{a}, a_0)$ y x_{n+1}^{ce} y además

$$x_{n+1}^{san} = x_n^{san} + \frac{a_0}{n^{\mathbf{a}}} (x_{n+1}^{ce} - x_n^{san}) = x_{n+1}^{ce} + \left(1 - \frac{a_0}{n^{\mathbf{a}}}\right) (x_n^{san} - x_{n+1}^{ce}) \quad (5.3)$$

y así cuando $0 < a_0/n^{\mathbf{a}} < 1$, $x_{n+1}^{san}(\mathbf{a}, a_0)$ varía desde $x_n^{san}(\mathbf{a}, a_0)$ hacia x_{n+1}^{ce} , mientras cuando $1 < a_0/n^{\mathbf{a}}$, $x_{n+1}^{san}(\mathbf{a}, a_0)$ se aleja de x_{n+1}^{ce} en la dirección opuesta. Aunque en principio podría parecer que tal proposición es demasiado restrictiva, puesto que C debe ser igual a R , no lo es tanto, pues a menudo los x generados según (5.1) no se dispersarán demasiado.

Notar además que la Expresión (5.3) tiene muchas similitudes con la dada en el Corolario 6.1 para los puntos seleccionados según el diseño *hibrido* introducido en el Capítulo 6. Estas expresiones podrían servir para construir otras familias de diseños utilizando los mismos pesos, pero cambiando los puntos de referencia a utilizar, tal como proponemos en la Sección 5.5 de este capítulo. Obsérvese por último que a través de (5.3) se podría extender estos diseños a problemas con más de una variable de control ($k > 1$).

El siguiente corolario expresa la posición relativa de $x_{n+1}^{san}(\mathbf{a}, a_0)$ con respecto a x_{n+1}^{ce} , $r_n(\mathbf{a}, a_0)$, en función de la posición de $x_n^{san}(\mathbf{a}, a_0)$. Este resultado nos ayudará a justificar el rango considerado para los parámetros \mathbf{a} y a_0 , y a interpretar de forma intuitiva cuales deben ser los mejores valores para dichos parámetros.

Corolario 5.1 Sea $x_{n+1}^{san}(\mathbf{a}, a_0)$ definido como en (5.1), con $\hat{y}_n(x_n) = f(x_n, \hat{\mathbf{b}}_n)$ y $\hat{f}'_n(\mathbf{q}; \mathbf{b}) = f'(x_n, \hat{\mathbf{b}}_n)$ y sea x_{n+1}^{ce} definido con el $\hat{y}_n(x)$ calculado a partir de los datos (X'_n, Y'_n) elegidos y observados a través de $d_{san}(\mathbf{a}, a_0)$ y $C=R$. Para el modelo lineal simple y $n = 1, 2, \dots, N - 2$,

$$r_n(\mathbf{a}, a_0) = \frac{x_{n+1}^{san}(\mathbf{a}, a_0) - x_{n+1}^{ce}}{x_n^{san}(\mathbf{a}, a_0) - x_{n+1}^{ce}} = 1 - \frac{a_0}{n^a}.$$

Su demostración es inmediata a partir de la Proposición 5.1, sin más que sumar y restar $x_n^{san}(\mathbf{a}, a_0)$ en el numerador de $r_n(\mathbf{a}, a_0)$. El valor de $r_n(\mathbf{a}, a_0)$ será siempre menor o igual a uno y como muestra la Figura 5.1, cuando $r_n(\mathbf{a}, a_0)$ es positivo $x_{n+1}^{san}(\mathbf{a}, a_0)$ pertenece al intervalo $(x_{n+1}^{ce}, x_n^{san}(\mathbf{a}, a_0))$ y cuando es más negativo se aleja de dicho intervalo en la dirección contraria respecto de $x_n^{san}(\mathbf{a}, a_0)$.

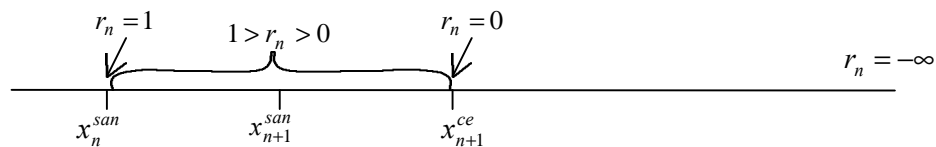


Figura 5.1 Muestra la situación de $x_{n+1}^{san}(\mathbf{a}, a_0)$ respecto a x_{n+1}^{ce} y $x_n^{san}(\mathbf{a}, a_0)$ según el valor de $r_n(\mathbf{a}, a_0) = 1 - a_0/n^a$.

El estudio de la función $r_n(\mathbf{a}, a_0)$, o del cociente a_0/n^a , nos permite entender la evolución de $x_{n+1}^{san}(\mathbf{a}, a_0)$ en función de n . Para ello se consideran cuatro posibilidades distintas atendiendo al signo de \mathbf{a} y el de $(a_0 - 1)$ con la idea de justificar la necesidad de $a_0 < 1$.

- a) Cuando $\mathbf{a} > 0$ y $a_0 > 1$, existe un n' tal que para todo $n \leq n'$, $a_0/n^{\mathbf{a}} > 1$. Para $n \leq n'$, $x_{n+1}^{san}(\mathbf{a}, a_0) \notin (x_n^{san}(\mathbf{a}, a_0), x_{n+1}^{ce})$, aunque $x_{n+1}^{san}(\mathbf{a}, a_0)$ se acerca a x_{n+1}^{ce} al aumentar n . Cuando $n > n'$, $x_{n+1}^{san}(\mathbf{a}, a_0)$ siempre cae en dicho intervalo y se acerca cada vez más a $x_n^{san}(\mathbf{a}, a_0)$.
- b) Cuando $\mathbf{a} > 0$ y $a_0 < 1$, para cualquier n , $a_0/n^{\mathbf{a}} < 1$ y $x_{n+1}^{san}(\mathbf{a}, a_0) \in (x_n^{san}(\mathbf{a}, a_0), x_{n+1}^{ce})$. Al aumentar n , $x_{n+1}^{san}(\mathbf{a}, a_0)$ se acerca a $x_n^{san}(\mathbf{a}, a_0)$.
- c) Cuando $\mathbf{a} < 0$ y $a_0 > 1$, para cualquier n , $a_0/n^{\mathbf{a}} > 1$ y $x_{n+1}^{san}(\mathbf{a}, a_0) \notin (x_n^{san}(\mathbf{a}, a_0), x_{n+1}^{ce})$. Al aumentar n , $x_{n+1}^{san}(\mathbf{a}, a_0)$ se aleja más y más de x_{n+1}^{ce} en el sentido contrario de $x_n^{san}(\mathbf{a}, a_0)$.
- d) Cuando $\mathbf{a} < 0$ y $a_0 < 1$, existe un n' , tal que para todo $n \leq n'$, $a_0/n^{\mathbf{a}} < 1$. Para $n \leq n'$, $x_{n+1}^{san}(\mathbf{a}, a_0) \in (x_n^{san}(\mathbf{a}, a_0), x_{n+1}^{ce})$ y la elección $x_{n+1}^{san}(\mathbf{a}, a_0)$ se acerca a x_{n+1}^{ce} al aumentar n . Cuando $n > n'$, $x_{n+1}^{san}(\mathbf{a}, a_0)$ se aleja de x_{n+1}^{ce} en el sentido contrario de $x_n^{san}(\mathbf{a}, a_0)$.

La opción c) no es deseable pues desde los primeros estadios $x_{n+1}^{san}(\mathbf{a}, a_0)$ se aleja cada vez más de x_{n+1}^{ce} al aumentar n , en contra del hecho de que el diseño *certainty equivalence* tiene un buen comportamiento asintótico.

En el caso de la opción a), en los estadios iniciales $r_n(\mathbf{a}, a_0) < 0$, con lo que $x_{n+1}^{san}(\mathbf{a}, a_0)$ queda situado fuera del intervalo $(x_n^{san}(\mathbf{a}, a_0), x_{n+1}^{ce})$, en el lado opuesto al punto $x_n^{san}(\mathbf{a}, a_0)$. Esto hace que en esos primeros estadios los puntos de experimentación queden demasiado dispersos y tengan una contribución demasiado grande sobre la pérdida total. Esto se ilustra en la Figura 5.2, donde se representan los tres primeros puntos que selecciona el diseño *san* bajo la opción a) apreciándose la gran distancia entre $x_2^{san}(\mathbf{a}, a_0)$ y $x_3^{san}(\mathbf{a}, a_0)$.

Para b) y d), a_0 es negativo y el diseño *san* observa lejos de la elección *certainty equivalence* secuencial en la dirección del último punto observado. En el límite, para $a_0 = -\infty$, el diseño *san* nos lleva a observar siempre en puntos cercanos a la frontera de C , tal como lo harían los diseños *D-óptimos* y *G-óptimos* propuestos para contextos en los que lo que importa es la estimación de \mathbf{q} .

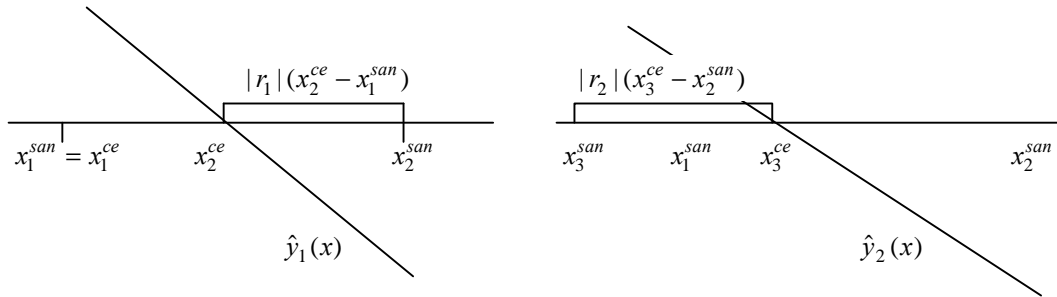


Figura 5.2 Representación, en un caso hipotético, de los puntos seleccionados por el diseño *san* en los tres primeros estadios, si $a_0 > 1$ y $\mathbf{a} > 0$.

Parece lógico pensar que el mejor diseño de entre las *san* sería aquel que en las primeras etapas se alejara un poco de *certainty equivalence* para mejorar las elecciones *certainty equivalence* posteriores y así al acercarse a las mismas garantizar un buen control. Bajo estas consideraciones, la opción d) con $a_0 \in [0,1]$ y $\mathbf{a} < 0$, parece la más acertada. Experimentalmente se ha comprobado que para los parámetros óptimos a_0^* y \mathbf{a}^* , n' verifica que $a_0^* = (n')^{\mathbf{a}^*}$ con $n' \in (N/2, N-1]$. La conclusión es que es aconsejable escoger valores de a_0 y de \mathbf{a} tales que en los estadios iniciales se verifique que $0 < a_0/n^{\mathbf{a}} < 1$.

5.2.3 Búsqueda del mejor diseño *san*.

Para encontrar el diseño $d_{san}(\mathbf{a}^*, a_0^*)$ que minimiza $r_i(\mathbf{a}, a_0) = r_i(d_{san}(\mathbf{a}, a_0))$ se procederá de manera análoga a las descritas en las Secciones 3.2.2 para los diseños *Fieller* y 4.2 para *HPD*. Para cada (\mathbf{a}, a_0) se simula repetidamente \mathbf{b} a partir de $\mathbf{p}_0(\mathbf{b})$ y se usa $d_{san}(\mathbf{a}, a_0)$ sobre datos simulados de acuerdo con el modelo para $(y|x^{san}, \mathbf{b})$, estimando $r_i(\mathbf{a}, a_0)$, $i=1,2,3$, a través del promedio de las pérdidas observadas. Buscando entre los distintos $d_{san}(\mathbf{a}, a_0)$ se encontrará el $d_{san}(\mathbf{a}^*, a_0^*)$ que sea el mejor para el problema, adaptando esta familia de diseños a la pérdida y al modelo considerado.

Se sabe que $d_{san}(\mathbf{a}^*, a_0^*)$ siempre mejora a los diseños *certainty equivalence* secuencial y no secuencial. Cuando (\mathbf{a}^*, a_0^*) está cerca de $(0,1)$ o de $(\mathbf{a}, 0)$, los diseños *certainty equivalence* secuencial y no secuencial, respectivamente, serán los “mejores” en esta familia. A continuación se estudia el comportamiento de las pérdidas esperadas, $r_i(\mathbf{a}, a_0)$, en función de \mathbf{a} y a_0 , para el modelo de regresión simple con dos prioris concretas. Después se realiza un estudio de sensibilidad en el que se muestra como \mathbf{a}^* y a_0^* dependen de la priori escogida.

5.3 Aplicación al modelo lineal.

Para ilustrar el uso de los diseños *san*, se explora su utilización en el modelo lineal normal con $f(x; \mathbf{b}) = \mathbf{b}_0 + \mathbf{b}_1 x$ para las mismas prioris que en la Sección 3.3. Primero se asume una distribución a priori tal que $\mathbf{b}_0 \sim N(0, \mathbf{s}_{\mathbf{b}_0}^2 = .4^2)$ y $\mathbf{b}_1 \sim N(1, \mathbf{s}_{\mathbf{b}_1}^2 = 2.^2)$ con \mathbf{b}_0 independiente de \mathbf{b}_1 . En este caso la superficie es estimada por $\hat{y}_n(x) = E_n[\mathbf{b}_0] + E_n[\mathbf{b}_1]x_n$ y la derivada por $\hat{f}'_n(\mathbf{q}; \mathbf{b}) = E_n[\mathbf{b}_1]$.

La primera columna de la Figura 5.3 muestra los gráficos de contorno para dicha priori de las $r_i(\mathbf{a}, a_0)$ estimadas para las tres funciones de pérdida descritas en la Sección 1 del Capítulo 1, con $N = 25$, $S = 100$ y $T = 0$. Recuérdese que $i = 1$ corresponde a la suma de desviaciones en valor absoluto, $i = 2$ a desviaciones cuadráticas e $i = 3$ a la pérdida con dos etapas. Estos contornos son construidos a partir de los valores de $r_i(\mathbf{a}, a_0)$ evaluados para una rejilla igualmente espaciada de 7×8 valores de (\mathbf{a}, a_0) . Para los gráficos de esta primera columna $\mathbf{s}^2 = .2^2$ y $C = [-1, 1]$.

El comportamiento de los gráficos de contorno para las tres pérdidas es muy similar así que sólo se comentará el caso de la pérdida cuadrática. Para esta priori, $\hat{r}_2(d_{ce}) =$

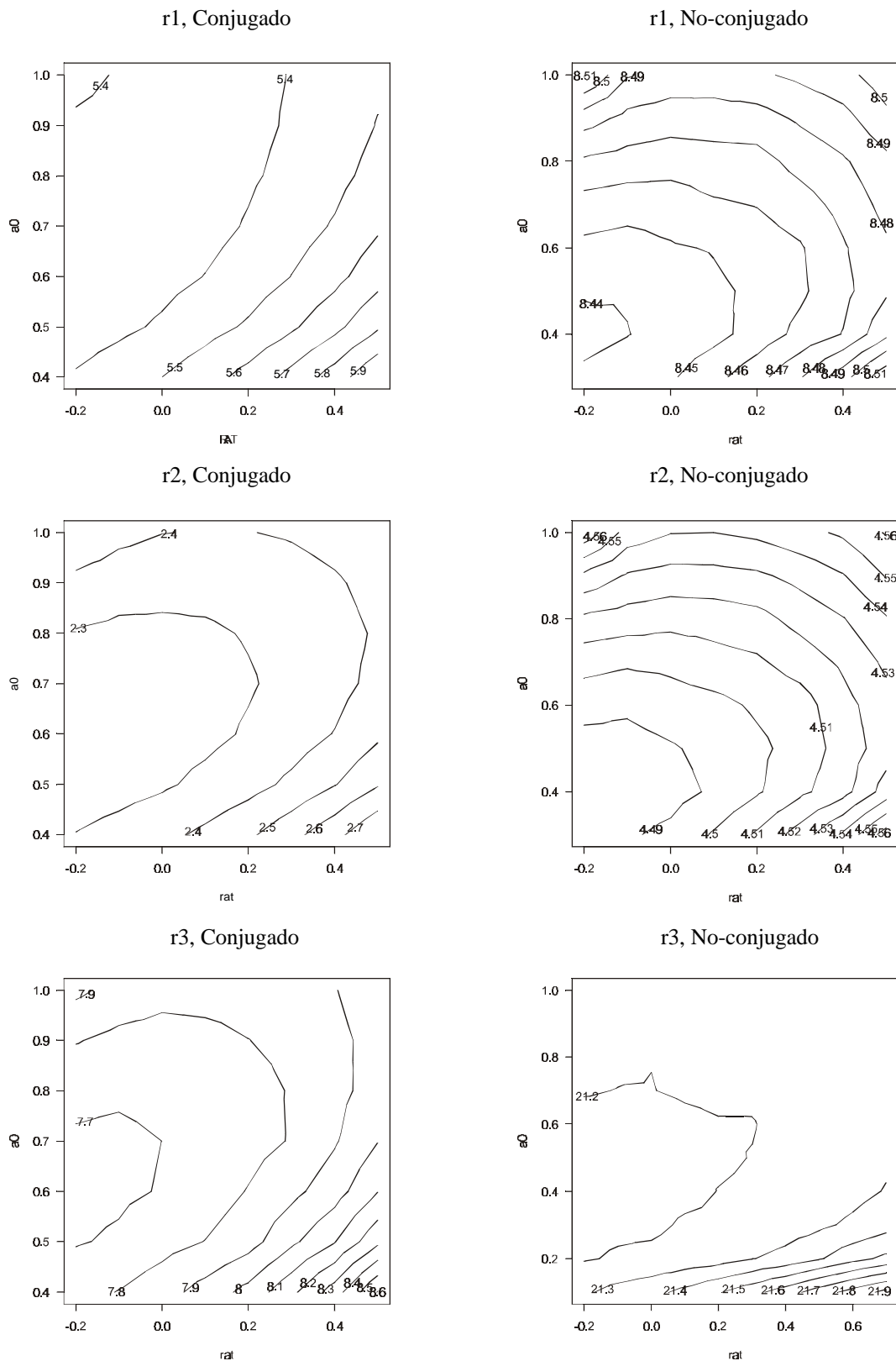


Figura 5.3. Gráfico de contorno de $\hat{r}_i(\mathbf{a}, a_0)$, $i=1,2,3$, con $N=25$, $S=100$, $T=0$ y $C=[-1,1]$. La primera columna es para la priori $\mathbf{b}_0 \sim N(\mathbf{m}_{b_0} = 0., \mathbf{s}_{b_0}^2 = .4^2)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1} = 1., \mathbf{s}_{b_1}^2 = 2.^2)$, \mathbf{b}_0 independiente de \mathbf{b}_1 y $\mathbf{s}^2 = .2^2$. La segunda columna es para $\mathbf{q} \sim U(-.5, .5)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1} = 1., \mathbf{s}_{b_1}^2 = 2.^2)$, \mathbf{b}_1 independiente de \mathbf{q} y $\mathbf{s}^2 = .4^2$.

$\hat{r}_2(\mathbf{a} = 0., a_0 = 1.) = 2.4051$ con un error estándar estimado de .0011, un valor típico sobre el rango de (\mathbf{a}, a_0) estudiado. Se tiene que $(\hat{\mathbf{a}}^{2*}, \hat{a}_0^{2*}) = (-.2, .6)$, y que $\hat{r}_2(\hat{\mathbf{a}}^{2*} = -.2, \hat{a}_0^{2*} = .6) = 2.2212$ con un error estándar de .00056. Para las tres pérdidas $\hat{\mathbf{a}}^{i*} < 0$ y $a_0^{i*} < 1$ y, por tanto, corresponden a la opción d) descrita en la Sección 5.2.2.

También se ilustra en el supuesto de que $y_n | x_n, \mathbf{b} \sim N(\mathbf{b}_1(x - \mathbf{q}), \mathbf{s}^2)$, con una priori no conjugada tal que $\mathbf{q} \sim U(-.5, .5)$, $\mathbf{b}_1 \sim N(1., \mathbf{s}_{b_1}^2 = .2^2)$ y \mathbf{b}_1 independiente de \mathbf{q} . En este último caso la superficie será estimada por $\hat{y}_n(x) = \mathbf{b}_{1n}^{pmo}(x_n - \mathbf{q}_n^{pmo})$ y la derivada por $\hat{f}'_n(\mathbf{q}; \mathbf{b}) = \mathbf{b}_{1n}^{pmo}$, tal como se detalla en la Sección 1 del Apéndice II.

Los gráficos de contorno de las $r_i(\mathbf{a}, a_0)$ presentados en la segunda columna de la Figura 5.3 corresponden a esta segunda priori con $\mathbf{s}^2 = .4^2$. Aquí el comportamiento para las tres pérdidas sigue siendo muy parecido con lo que sólo se tratará el de la segunda. Para esta priori, $\hat{r}_2(d_{ce}) = \hat{r}_2(\mathbf{a} = 0., a_0 = 1.) = 4.5401$, con un error estándar estimado de ,00019 y $(\hat{\mathbf{a}}^{2*}, \hat{a}_0^{2*}) = (-.3, .4)$ con $\hat{r}_2(\hat{\mathbf{a}}^{2*} = -.3, \hat{a}_0^{2*} = .4) = 4.4817$ y un error estándar de .00021.

5.4 Análisis de sensibilidad.

Hasta aquí hemos ilustrado cómo actúan los diseños para una priori determinada. En esta sección se analiza el comportamiento del mejor diseño *san* en función de la priori, N y C . Así como en la Sección 5.3 hemos considerado tres pérdidas distintas, en ésta restringiremos el estudio al caso de $r_2(d)$.

Para evaluar la mejora relativa de $d_{san}(\mathbf{a}^*, a_0^*)$ con respecto de cualquier otro $d_{san}(\mathbf{a}, a_0)$, se define $rr(\mathbf{a}, a_0)$ como la diferencia relativa entre la pérdida esperada bajo $d_{san}(\mathbf{a}^*, a_0^*)$ y la pérdida esperada bajo $d_{san}(\mathbf{a}, a_0)$,

$$rr(\mathbf{a}, a_0) = \frac{r_2(\mathbf{a}, a_0) - r_2(\mathbf{a}^*, a_0^*)}{r_2(\mathbf{a}^*, a_0^*)}.$$

Notar que la mejora relativa de $d_{san}(\mathbf{a}^*, a_0^*)$ sobre el diseño *certainty equivalence* es $rr(\mathbf{a}=0, a_0=1)$.

El modelo escogido para el análisis es el lineal normal con $f(x; \mathbf{b}) = \mathbf{b}_0 + \mathbf{b}_1 x$ y con la misma priori $\mathbf{p}_0(\mathbf{b})$ que en la Sección 3.4, es decir, que $\mathbf{b}_0 \sim N(0, \mathbf{s}_{b_0}^2 = .5^2)$, $\mathbf{b}_1 \sim N(1, \mathbf{s}_{b_1}^2)$ con \mathbf{b}_1 independiente de \mathbf{b}_0 , $T=0$ y $C=[-R, R]$. Como en la Sección 3.4, para evaluar $r_2(\mathbf{a}, a_0)$ se simulan 3.000.000 de $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1)$ distribuidas como $\mathbf{p}_0(\mathbf{b})$, y para cada \mathbf{b} se ejecuta el diseño $d_{san}(\mathbf{a}, a_0)$ sobre las $y_i(x_i^{san})$'s simuladas usando la distribución $(y_i | x_i^{san}, \mathbf{b})$. El valor de $r_2(\mathbf{a}, a_0)$ se estima a partir del promedio de las pérdidas observadas. Como la desviación estándar estimada de $\hat{r}_2(\mathbf{a}, a_0)$ es del orden de .0007, el valor (\mathbf{a}^*, a_0^*) que minimiza $r_2(\mathbf{a}, a_0)$ puede ser obtenido de forma muy precisa utilizando el procedimiento dado por Nelder y Mead (1965).

La Figura 5.4 explora la dependencia de a_0^* , \mathbf{a}^* y $rr(\mathbf{a}=0, a_0=1)$ respecto de \mathbf{s}_{b_1} , de \mathbf{s} , del horizonte N y de R . Cuando $N=20$ y $R=10$, a_0^* disminuye cuando aumenta \mathbf{s}_{b_1} y \mathbf{s} ; si tanto \mathbf{s} y \mathbf{s}_{b_1} son grandes, a_0^* está cerca de 0 y el diseño *certainty equivalence* secuencial es el mejor en la familia. Para \mathbf{s} pequeña y \mathbf{s}_{b_1} grande, el diseño *certainty equivalence* secuencial es el mejor. Notar que para todos estos ejemplos $0 < a_0^*$ y $-1 < \mathbf{a}^* < 1$.

Dado un R , cuando N aumenta, a_0^* tiende hacia 1 y el mejor diseño en esta familia estará cerca del *certainty equivalence*, como era de esperar ya que los diseños *certainty equivalence* son asintóticamente óptimos. Notar también que para R pequeño y N grande \mathbf{a}^* es negativo.

La Figura 5.4 también muestra que $rr(0,1)$ aumenta cuando lo hace \mathbf{s}_{b_1} , mientras es bastante insensible a los cambios en \mathbf{s} . Cuanto mayor es N , menor es la mejora de $d_{san}(\mathbf{a}^*, a_0^*)$ sobre el diseño *certainty equivalence*. Para regiones experimentales muy grandes, *certainty*

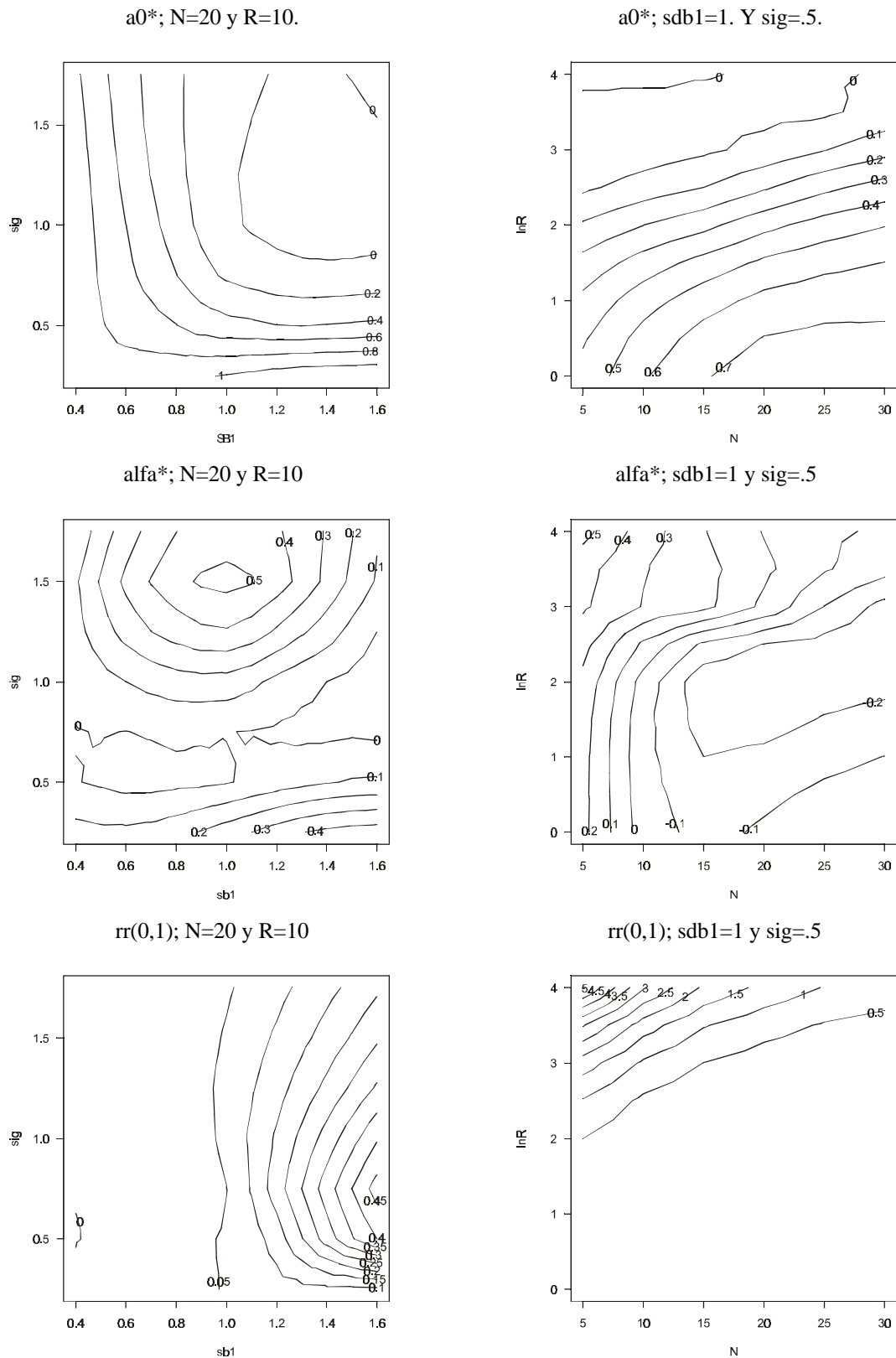


Figura 5.4. Gráfico de contorno de a_0^* , α^* y de la mejora relativa de $d_{san}(\mathbf{a}^*, a_0^*)$ sobre c.e., $rr(0,1)$, como una función de \mathbf{s} , \mathbf{s}_{b_1} , N y $\ln(R)$. $(y|x, \mathbf{b}) \sim N(\mathbf{b}_0 + \mathbf{b}_1 x, \mathbf{s}^2)$ con $\mathbf{b}_0 \sim N(\mathbf{m}_{b_0} = 0., \mathbf{s}_{b_0}^2 = .5^2)$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1} = 1., \mathbf{s}_{b_1}^2)$, \mathbf{b}_1 independiente de \mathbf{b}_0 , $C = [-R, R]$ y $T = 0$.

equivalence es extremadamente ineficiente y la mejora alcanzada a través de $d_{san}(\mathbf{a}^*, a_0^*)$ puede ser grande. Observar que para las prioris, horizontes y regiones experimentales utilizadas, las mejoras alcanzadas a través de $d_F(K^*, \mathbf{g}^*)$ son ligeramente mayores que las obtenidas a través de $d_{san}(\mathbf{a}^*, a_0^*)$.

5.6 Extensiones y comentarios.

Al igual que ocurría con los diseños *Fieller* definidos en el Capítulo 3, los diseños *san* introducen una acción de sondeo entorno a la elección *certainty equivalence* con el objetivo de mejorar los estimadores intermedios de los parámetros \mathbf{b} y así alcanzar un mejor control futuro. En los ejemplos presentados la computación requerida para obtener $(\hat{\mathbf{a}}^*, \hat{a}_0^*)$ no es demasiado intensiva, pero las mejoras relativas obtenidas respecto a *certainty equivalence* dependen del horizonte, del tamaño de la región experimental y de la priori, esto es, dependerá de cada situación en particular. En algunas circunstancias, podría ser útil adaptar aún más $d_{san}(\mathbf{a}, a_0)$ al problema, recalculando $(\hat{\mathbf{a}}_m^*, \hat{a}_{0m}^*)$ en algunos estadios intermedios usando la distribución a posteriori de \mathbf{b} en ese estadio. Estos cálculos requieren coste computacional adicional considerable y cuando lo hemos implementado las mejoras han sido ínfimas respecto a $d_{san}(\mathbf{a}^*, a_0^*)$.

En ausencia de formas de implementar los diseños óptimos, algunos métodos de optimización estocástica generarán conjuntos de diseños que tienen miembros con pérdida esperada pequeña. La simulación de MonteCarlo siempre será una opción a utilizar para buscar el mejor diseño dentro de dichos conjuntos. Por ejemplo, el problema de control multiperiodo puede reformularse como el que busca diseños d que maximizan $E[-\sum_{n=1}^N (y_n(x_n^d) - T)^2]$, y

adaptar los diseños definidos en Ginebra y Clayton (1995) a éstos. Observan y_{n+1} en el $x_{n+1}^{hb}(B) \in C$ más cercano al x maximizando $UB_n(x; B) = \hat{y}_n(x) + B s.d.(\hat{y}_n(x))$. Cuando $B = 0$, se obtiene el correspondiente diseño *certainty equivalence*, y cuanto mayor B el diseño favorecerá experimentar en puntos de C donde la información que se tiene de $y(x)$ es menor. Algunas funciones de pérdida permitirán a las familias de diseños además adaptarse al problema tratado. Si la pérdida es asimétrica con respecto a T , se puede extender la familia de diseños *san* al caso en que la elección de a_0 dependa del signo de $(\hat{y}_n - T)$.

Como se estableció en la Sección 5.2.2, los puntos del diseño *san* para el modelo lineal simple satisfacen la relación

$$x_{n+1}^{san} = x_{n+1}^{ce} + \left(1 - \frac{a_0}{n^a}\right)(x_n^{san} - x_{n+1}^{ce}). \quad (5.3)$$

Inmediatamente esto ya sugiere generalizaciones del diseño *san* para $k > 1$. Otra posibilidad es definir por analogía con (5.3) nuevos diseños construidos tomando como referencia elecciones distintas a la *certainty equivalence*.

En el Capítulo 6 se comprobará que el diseño *miope* es a menudo más eficiente que el *certainty equivalence*. Elijiendo al *miope* como diseño de referencia, una propuesta alternativa será experimentar en la secuencia de puntos generada a través de

$$x_{n+1} = x_{n+1}^{my} + \left(1 - \frac{a_0}{n^a}\right)(x_n - x_{n+1}^{my}). \quad (5.4)$$

Además (5.3) mejorará poco para N grande pues el diseño *certainty equivalence* es un miembro de la familia y (5.4) mejorará poco para N pequeño pues el diseño *miope* es miembro de ella, y recíprocamente.

Por último notar que los diseños que se presentan en el Capítulo 6 se pueden representar de la forma

$$x_{n+1} = x_{n+1}^{my} + \mathbf{I}_n(X, Y)(x_n^{ce} - x_{n+1}^{my}),$$

es decir, se construirán adoptando a los dos diseños a la vez como referencia (ver Corolario 6.1).

Apéndice 5.A Demostración Proposición 5.1.

La primera igualdad en (5.2) se sigue de sustituir $\hat{y}_n(x_n) = f(x_n; \hat{\mathbf{b}}_n)$ y $\hat{f}'_n(\mathbf{q}; \mathbf{b}) = \hat{\mathbf{b}}_{1n}$ en (5.1). La segunda igualdad es probada por inducción sobre n . Para $n = 1$,

$$x_2^{san}(\mathbf{a}, a_0) = \frac{1^{\mathbf{a}} - a_0}{1^{\mathbf{a}}} x_1^{ce} + \frac{a_0}{1^{\mathbf{a}}} x_2^{ce} = \sum_{i=1}^2 h_i(\mathbf{a}, a_0, i) x_i^{ce}.$$

Asumiendo (5.2) cierta para $n = r - 1$, entonces

$$x_{r+1}^{san}(\mathbf{a}, a_0) = \frac{(r^{\mathbf{a}} - a_0) \sum_{i=1}^r h_{r-1}(\mathbf{a}, a_0, i) x_i^{ce} + a_0 x_{r+1}^{ce}}{r^{\mathbf{a}}} = \sum_{i=1}^{r+1} h_r(\mathbf{a}, a_0, i) x_i^{ce},$$

donde $h_r(\mathbf{a}, a_0, i) = (r^{\mathbf{a}} - a_0) / (r^{\mathbf{a}}) h_{r-1}(\mathbf{a}, a_0, i)$, $i = 1, 2, \dots, r$, y $h_r(\mathbf{a}, a_0, i = r + 1) = a_0 / r^{\mathbf{a}}$.

Sustituyendo recursivamente h_{r-1} en función de h_{r-2} , se llega al resultado final. La demostración de que $\sum_{i=1}^{n+1} h_n(\mathbf{a}, a_0, i) = 1$ también se sigue por inducción sobre n , ya que se verifica para $n = 1$, y

$$\sum_{i=1}^{r+1} h_r(\mathbf{a}, a_0, i) = \frac{r^{\mathbf{a}} - a_0}{r^{\mathbf{a}}} \sum_{i=1}^r h_{r-1}(\mathbf{a}, a_0, i) + \frac{a_0}{r^{\mathbf{a}}} = 1. \blacksquare$$

El siguiente corolario muestra como se puede expresar el $x_{n+1}^{san}(\mathbf{a}, a_0)$ cuando se utiliza el modelo lineal simple y como estimadores de los parámetros los mínimos cuadrados. En general, será válido en los casos en que el *certainty equivalence* se pueda expresar como $x_{n+1}^{ce} = \bar{x}_n - \bar{y}_n / \hat{\mathbf{b}}_{1n}$.

Corolario 5.2 Sea $x_{n+1}^{san}(\mathbf{a}, a_0)$ definido como en (5.1), con el modelo $y = f(x; \mathbf{b}) + \mathbf{e} = \mathbf{b}_0 + \mathbf{b}_1 x + \mathbf{e}$ y con $\hat{y}_n(x_n) = f(x_n; \hat{\mathbf{b}}_n^{ls})$, $\hat{f}'_n(x; \mathbf{b}) = \hat{\mathbf{b}}_{1n}^{ls}$ y $x_{n+1}^{ce} = -\hat{\mathbf{b}}_{0n}^{ls} / \hat{\mathbf{b}}_{1n}^{ls}$, entonces

$$x_{n+1}^{san}(\mathbf{a}, a_0) = \bar{x}_n - \sum_{i=1}^n c_n(\mathbf{a}, a_0, i) \frac{\bar{y}_i}{\hat{\mathbf{b}}_{1i}^{ls}},$$

donde

$$c_n(\mathbf{a}, a_0, n) = \frac{a_0}{n^{\mathbf{a}}},$$

$$c_n(\mathbf{a}, a_0, i) = \frac{(n-1)(n^{\mathbf{a}} - a_0)}{n^{\mathbf{a}+1}} c_{n-1}(\mathbf{a}, a_0, i), \quad \text{para } i=1, \dots, n-1.$$

Demostración.

Se procederá por inducción sobre n en (5.1) haciendo uso de que $x_{n+1}^{ce} = -\hat{\mathbf{b}}_{0n}^{ls} / \hat{\mathbf{b}}_{1n}^{ls} = \bar{x}_n - \bar{y}_n / \hat{\mathbf{b}}_{1n}^{ls}$. Para $n=1$, se tiene que

$$x_2^{san}(\mathbf{a}, a_0) = \frac{(1^{\mathbf{a}} - a_0)x_1^{san}(\mathbf{a}, a_0) + a_0 x_2^{ce}}{1^{\mathbf{a}}} = \bar{x}_1 - \frac{a_0}{1^{\mathbf{a}}} \frac{\bar{y}_1}{\hat{\mathbf{b}}_{11}^{ls}}.$$

Asumiéndolo cierto hasta $n=r-1$, y teniendo en cuenta que se verifica siempre que $\bar{x}_n = ((n+1)\bar{x}_{n+1} - x_{n+1})/n$. La hipótesis de inducción queda como

$$x_r^{san}(\mathbf{a}, a_0) = \bar{x}_r - \frac{r-1}{r} \sum_{i=1}^{r-1} c_{r-1}(\mathbf{a}, a_0, i) \frac{\bar{y}_i}{\hat{\mathbf{b}}_{1i}^{ls}}.$$

De aquí, para $n=r$:

$$\begin{aligned} x_{r+1}^{san}(\mathbf{a}, a_0) &= \frac{(r^{\mathbf{a}} - a_0)x_r^{san}(\mathbf{a}, a_0) + a_0(\bar{x}_r - \bar{y}_r / \hat{\mathbf{b}}_{1r}^{ls})}{r^{\mathbf{a}}} \\ &= \bar{x}_r - \left[\frac{r-1}{r} \frac{r^{\mathbf{a}} - a_0}{r^{\mathbf{a}}} \sum_{i=1}^{r-1} c_{r-1}(\mathbf{a}, a_0, i) \frac{\bar{y}_i}{\hat{\mathbf{b}}_{1i}^{ls}} + \frac{a_0}{r^{\mathbf{a}}} \frac{\bar{y}_r}{\hat{\mathbf{b}}_{1r}^{ls}} \right], \end{aligned}$$

con lo que queda demostrado. ■

Nótese que la expresión de $c_n(\mathbf{a}, a_0, i)$ obtenida en este último corolario se relaciona con la expresión de $h_n(\mathbf{a}, a_0, i)$ dada en la Proposición 5.1 a través de

$$c_n(\mathbf{a}, a_0, i) = \frac{i a_0 [(i-1)!]^{\mathbf{a}} \prod_{j=i+1}^n (j^{\mathbf{a}} - a_0)}{(n!)^{\mathbf{a}}} = \frac{i}{n} h_n(\mathbf{a}, a_0, i+1), \quad \text{para } i=1, \dots, n.$$

Un problema de la utilización de estimadores que no usan información a priori como los mínimos cuadrados, es la necesidad de especificar al menos dos puntos iniciales para poder empezar a realizar el experimento. Según los valores iniciales escogidos, las pérdidas cometidas en los primeros estadios pueden sobrepasar en mucho las mejoras que se obtienen sobre *certainty equivalence*. Anderson y Taylor (1976) utilizan los estimadores mínimos cuadrados en problemas de control estudiando sus propiedades estadísticas y el efecto de la elección arbitraria de los dos puntos iniciales.

Parte III:

Múltiples variables de control

CAPITULO 6.

DISEÑOS *HÍBRIDOS*.

Aunque los diseños *certainty equivalence* pueden tener buenas propiedades asintóticas, como muestran Hu (1997, 98) y Chen y Hu (1998), no tienen en consideración la incertidumbre en la estimación de la superficie y su eficiencia puede ser muy baja para N pequeños y moderados. Por otro lado, los diseños *miopes* están próximos al óptimo para N pequeña, pero su eficacia degenera cuando ésta crece. En este capítulo se construye una familia de diseños, indexados por un único parámetro $H \in [0,1]$ y denotados por $d_{hb}(H)$, que contienen tanto a nuestra versión bayesiana del *certainty equivalence*, definida en el Capítulo 2, como a los diseños *miopes*. Estos diseños *híbridos* incorporan la necesidad del aprendizaje activo sobre \mathbf{b} para disminuir la desviación futura respecto a T .

A diferencia de los diseños introducidos en los capítulos anteriores, los *híbridos* sólo están indexados por un único parámetro H , frente a los dos necesarios para los diseños *San*, *Fieller* y los basados en regiones creíbles, lo que simplifica enormemente la búsqueda del mejor diseño en esta familia. Además, el soporte del índice H es un intervalo acotado, a diferencia del índice K de los diseños *Fieller*. Por otro lado, las familias de diseños *Fieller*, *San* y *HPD* son difícilmente extensibles a problemas con más de una variable de control, tanto por su complejidad computacional como por el aumento de parámetros a utilizar. Los diseños *híbridos*

en cambio permiten la generalización a más de una variable de control ($k > 1$) sin necesidad de usar más que un índice de diseño. Además tiene la ventaja que para el modelo lineal el espacio de búsqueda del punto de diseño es unidimensional para cualquier número de variables de control k . También hay que indicar que por la propia construcción, el diseño *miope* forma parte de dicha familia, hecho que no ocurre en el resto de los diseños introducidos en los capítulos anteriores y que permite aprovechar la eficiencia de los diseños *miopes* cuando el problema tiene un horizonte N relativamente pequeño.

En la Sección 6.1 se definen los diseños *híbridos*, en la Sección 6.2 se explora la actuación de tales diseños para el modelo lineal y en la Sección 6.3 se estudia el caso particular de un modelo lineal normal con priori conjugada. En la Sección 6.4 se realiza un análisis de sensibilidad que ilustra el comportamiento del mejor diseño *híbrido* y la mejora relativa con respecto a *certainty equivalence* y *miope*, dependiendo de la priori, el horizonte N y el tamaño de la región experimental, tanto para el caso de un único control como para más de una variable de control. En la Sección 6.5 se comentan posibles mejoras de los diseños *híbridos* y se discute la aplicación de estos diseños a modelos más complicados. El capítulo incluye apéndices con las demostraciones de los resultados presentados en el capítulo.

6.1 Definición de diseños *híbridos*.

Recordar que la elección *certainty equivalence* en el caso en el que se estima la superficie a través de $\hat{y}_n(x) = E_n[y(x)]$, se obtiene de resolver en x en cada estadio $E_n[y(x) - T] = 0$, que es equivalente a minimizar en x la función $(E_n[y(x) - T])^2$. Por otra parte, para la pérdida cuadrática, la elección *miope* minimiza en x para cada estadio n la función

$E_n[(y(x)-T)^2]$. Los diseños *híbridos* definidos a continuación combinan los dos criterios anteriores.

Definición 6.1 Dado $H \in (0,1]$, el diseño $d_{hb}(H)$ observa y_{n+1} en el punto $x_{n+1}^{hb}(H)$ minimizando en x

$$hb_{n+1}(x;H) = (1-H)(E_n[y(x)-T])^2 + H E_n[(y(x)-T)^2] =$$

$$(E_n[y(x)-T])^2 + H V_n[y(x)-T] = E_n[(y(x)-T)^2] + (H-1)V_n[y(x)-T].$$

Cuando $H=0$, $x_{n+1}^{hb}(H)$ es la solución de $E_n[y(x)-T]=0$ que minimiza $V_n[y(x)-T]$.

Notar que el diseño *híbrido* se define a través de la minimización de una combinación convexa entre $(E_n[y(x)-T])^2$ y $E_n[(y(x)-T)^2]$, que son precisamente los criterios que se minimizan para calcular las elecciones *certainty equivalence* y *miope* bayesianas. Tal como se desprende del Corolario 6.1, en el caso particular de un modelo lineal, el mínimo de esta combinación convexa es nuevamente una combinación convexa de las soluciones *certainty equivalence* y *miope*.

Como $x_{n+1}^{hb}(H=0) = x_{n+1}^{ce}$, el diseño $d_{hb}(H=0)$ coincide con nuestra elección *certainty equivalence* y, al ser $x_{n+1}^{hb}(H=1) = x_{n+1}^{my}$, $d_{hb}(H=1)$ reproduce el diseño *miope*. Cuando $0 < H < 1$, el diseño $d_{hb}(H)$ perturba la elección *miope* incentivando la observación en puntos $x_{n+1}^{hb}(H)$ donde la varianza de predicción, $V_n[y(x_{n+1}^{hb}(H))]$, es más grande que la de la elección *miope*, $V_n[x_{n+1}^{my}]$, y menor que la de la elección *certainty equivalence*, $V_n[x_{n+1}^{ce}]$, tal como se muestra en el Lema 6.3 incluido en el Apéndice 6.A.1. Al alejarse de la elección *miope*, los diseños *híbridos* se alejan también de los puntos para los que el valor previsto de y está más cerca de T , para aprender activamente sobre $E[y|x] = f(x; \mathbf{b})$ y así, alcanzar un mejor control futuro, de una manera análoga a las estrategias propuestas por Lai (1987) para el problema del *bandit* de dos brazos y los diseños de Ginebra y Clayton (1995) para el *Response Surface Bandit*.

Para encontrar el diseño *híbrido* que minimice

$$r_2(H) = r_2(d_{hb}(H)) = E_0[\sum_{n=1}^N (y_n(x_n^{hb}(H)) - T)^2],$$

denotado por $d_{hb}(H^*)$, se procede a través de simulación de MonteCarlo sobre la distribución conjunta de los parámetros y del espacio muestral como sigue: para cada H se simulan repetidamente valores de \mathbf{b} de acuerdo con la distribución $\mathbf{p}_0(\mathbf{b})$, y para cada uno de estos \mathbf{b} , se simula el uso del diseño $d_{hb}(H)$ sobre datos procedentes del modelo para $y|\mathbf{b}$, estimando su pérdida esperada, $r_2(H)$, a través de la media de las pérdidas observadas. Buscando entre los distintos $d_{hb}(H)$, se encuentra el $d_{hb}(H^*)$ que se estima que es el “mejor” diseño *híbrido* para la función de pérdida, la priori sobre \mathbf{b} y el modelo especificado.

Sabemos que $d_{hb}(H^*)$ siempre mejora tanto el diseño *certainty equivalence* como el *miope*. De hecho, cuando H^* está cercano a 0, el diseño *certainty equivalence* secuencial será el mejor diseño *híbrido* y cuando H^* está cercano a 1, el mejor diseño *híbrido* será el *miope*. Así, $d_{hb}(H^*)$ ofrece un buen compromiso entre el diseño que sería optimal si el estimador de la superficie fuera la superficie (con buenas propiedades asintóticas) y el diseño que sería optimal si sólo quedara por tomar una única observación (que es eficiente para N pequeña). De alguna manera, H^* puede ser entendido como una medida del grado de miopicidad requerido por el problema planteado.

El siguiente lema facilita la computación del $hb_{n+1}(x; H)$ cuando los parámetros \mathbf{b} son independientes del error \mathbf{e} .

Lema 6.1 Cuando \mathbf{e} es independiente de \mathbf{b} , $E[y|x, \mathbf{b}] = f(x; \mathbf{b})$ y $V[\mathbf{e} | x, \mathbf{b}] = \mathbf{s}^2$,

$$hb_{n+1}(x; H) = (E_n[f(x; \mathbf{b}) - T])^2 + H V_n[f(x; \mathbf{b}) - T] + \hat{\mathbf{S}}_n^2 H$$

con $\hat{\mathbf{S}}_n^2$ una estimación de la varianza en el estadio n .

Para modelos en los que la computación y minimización de $hb_{n+1}(x; H)$ es demasiado complicada, una familia alternativa de diseños que aproximan a los diseños *híbridos* se obtiene experimentando en los x_n 's que minimizan un estimador de $hb_{n+1}(x; H)$ como puede ser $\hat{hb}_{n+1}(x; H) = (\hat{y}_n(x) - T)^2 + H \text{vâr}[\hat{y}_n(x) - T]$. En un primer momento nos planteamos la posibilidad de definir diseños que maximicen $UB_n(x; \mathbf{b}) = -(\hat{y}_n(x) - T)^2 - B.s.d.((y_n(x) - T)^2)$, por ser las extensiones naturales de los diseños propuestos por Ginebra y Clayton (1995) para el *Response Surface Bandit*, pero de momento descartamos su implementación dada la complejidad de cálculo de $UB_n(x; \mathbf{b})$. A continuación se calcula $x_{n+1}^{hb}(H)$ para el modelo lineal.

6.2 Diseños *híbridos* para el modelo lineal.

6.2.1 Modelo lineal de primer orden.

Para ilustrar el uso de los diseños *híbridos*, se explora su utilización en el modelo de regresión lineal de primer orden para el que $x_n^{hb}(H) = (x_{n1}(H), \dots, x_{nk}(H))$ siempre existe, es único y puede ser obtenido de forma cerrada en función de los dos primeros momentos de la distribución a posteriori de \mathbf{b} . La siguiente proposición muestra la ecuación del $x_{n+1}^{hb}(H)$ para estos modelos.

Proposición 6.1 Sean $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i$, $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$, y $(\mathbf{b}; \mathbf{s}^2) \sim \mathbf{p}_0(\mathbf{b}, \mathbf{s}^2)$, con $\mathbf{b}' = (\mathbf{b}_0, \dots, \mathbf{b}_k)$ independiente de \mathbf{e} . Cuando $H \in [0, 1]$, el diseño $d_{hb}(H)$ observa y_{n+1} en

$$x_{n+1}^{hb}(H) = V_n^{-1}[\mathbf{b}_I] \left(E_n[\mathbf{b}_I] \frac{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] COV_n[\mathbf{b}_0, \mathbf{b}_I] + (T - E_n[\mathbf{b}_0])}{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I] + H} - COV_n[\mathbf{b}_0, \mathbf{b}_I] \right)$$

donde $\mathbf{b}'_I = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k)$ es el vector de todos los parámetros excepto \mathbf{b}_0 y $COV_n[\mathbf{b}_0, \mathbf{b}_I]$ es el vector $k \times 1$ de covarianzas a posteriori entre \mathbf{b}_0 y los k elementos de \mathbf{b}_I .

Notar que esta proposición no asume ninguna distribución sobre \mathbf{e} y que \mathbf{s}^2 no tiene porque ser conocida, aunque sí independiente de x . El Lema 6.5 del Apéndice 6.A.2 presenta una formulación alternativa de $x_{n+1}^{hb}(H)$ cuando $H \neq 0$. Como una consecuencia inmediata de la Proposición 6.1 se tiene el siguiente corolario.

Corolario 6.1 Sean $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i$, $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$ y $(\mathbf{b} : \mathbf{s}^2) \sim \mathbf{p}_0(\mathbf{b}, \mathbf{s}^2)$ con $\mathbf{b}' = (\mathbf{b}_0, \dots, \mathbf{b}_k)$ independiente de \mathbf{e} y $\mathbf{b}'_I = (\mathbf{b}_1, \dots, \mathbf{b}_k)$. Entonces

$$x_{n+1}^{hb}(H) = x_{n+1}^{my} + \frac{(1-H)E_n[\mathbf{b}_I]'V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I]}{H + E_n[\mathbf{b}_I]'V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I]}(x_{n+1}^{ce} - x_{n+1}^{my}) =$$

$$x_{n+1}^{ce} + \frac{(1 + E_n[\mathbf{b}_I]'V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I])H}{H + E_n[\mathbf{b}_I]'V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I]}(x_{n+1}^{my} - x_{n+1}^{ce}).$$

Por tanto, para los modelos lineales de primer orden (sin interacciones ni términos cuadráticos), el conjunto de todos los puntos $x_{n+1}^{hb}(H)$ obtenidos al variar H es el segmento de línea recta que une x_{n+1}^{ce} con x_{n+1}^{my} . Del Corolario 6.1, también se sigue que cuando H varía de 0 a 1, $x_{n+1}^{hb}(H)$ se mueve de forma monótona desde la elección *certainty equivalence* a la elección *miope*, sin más que comprobar que la derivada de $x_{n+1}^{hb}(H)$ con respecto a H

$$\frac{\partial x_{n+1}^{hb}(H)}{\partial H} = - \frac{E_n[\mathbf{b}_I]'V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I](1 + E_n[\mathbf{b}_I]'V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I])}{(H + E_n[\mathbf{b}_I]'V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I])^2}(x_{n+1}^{ce} - x_{n+1}^{my})$$

tiene signo constante y, por tanto, cada H define de forma unívoca un diseño.

En la Proposición 6.2 se prueba además que para el modelo lineal normal con priori conjugada esta línea recta no depende de las $(x_i^d, y_i(x_i^d))$ seleccionadas y observadas hasta la

etapa n , X_n^d e Y_n , y por ello se trata de la misma recta para cualquier estadio n . El siguiente corolario, muestra la expresión de $x_{n+1}^{hb}(H)$ para el modelo de regresión lineal simple, $k=1$.

Corolario 6.2 Sean $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \mathbf{b}_1 x$, $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$ y $(\mathbf{b} : \mathbf{s}^2) \sim \mathbf{p}_0(\mathbf{b}, \mathbf{s}^2)$ con \mathbf{b} independiente de \mathbf{e} . Cuando $H \in [0, 1]$, el diseño $d_{hb}(H)$ observa y_{n+1} en

$$x_{n+1}^{hb}(H) = -\frac{(E_n[\mathbf{b}_0] - T) E_n[\mathbf{b}_1] + H \text{Cov}_n[\mathbf{b}_0, \mathbf{b}_1]}{E_n^2[\mathbf{b}_1] + H V_n[\mathbf{b}_1]} =$$

$$-\frac{(1-H) E_n[\mathbf{b}_0] E_n[\mathbf{b}_1] + H E_n[\mathbf{b}_0 \mathbf{b}_1] - T E_n[\mathbf{b}_1]}{(1-H) E_n^2[\mathbf{b}_1] + H E_n[\mathbf{b}_1^2]}.$$

Su demostración sigue de forma inmediata de la simplificación de la expresión del $x_{n+1}^{hb}(H)$ dada en la Proposición 6.1. En particular, para los modelos lineales simples se tiene que $x_{n+1}^{ce} = x_{n+1}^{hb}(0) = (T - E_n[\mathbf{b}_0]) / E_n[\mathbf{b}_1]$ y $x_{n+1}^{my} = x_{n+1}^{hb}(1) = (T E_n[\mathbf{b}_1] - E_n[\mathbf{b}_0 \mathbf{b}_1]) / E_n[\mathbf{b}_1^2]$.

Es usual que el valor de las variables de control x esté restringido a pertenecer a una región $C \subset R^k$. Cuando el punto que minimiza $hb_{n+1}(H)$ no se encuentra dentro de dicha región habrá que realizar una minimización con restricciones. En los ejemplos que se exponen, y sin pérdida de generalidad, consideramos regiones de tipo esféricas de radio R centradas en el origen. Dado que la expresión del $hb_{n+1}(H)$ para los modelos lineales son formas cuadráticas, cuando el punto que minimiza $hb_{n+1}(H)$ no pertenece a C los puntos a buscar caerán en la frontera de dicha región. La resolución de tal problema lleva a emplear el método iterativo de Newton-Raphson para sistemas de ecuaciones no lineales. El Apéndice 6.A.4 muestra el sistema no lineal a resolver para el cálculo del punto de la frontera y presenta la matriz jacobiana resultante para tal método. Además se propone un punto de inicio para dicho método iterativo.

6.2.2 Modelo lineal con priori conjugada (\mathbf{s}^2 conocida).

Un caso particular importante es el del modelo lineal normal con prioris normales sobre los parámetros \mathbf{b} . Para estos modelos, y por disponer de los momentos a posteriori de los parámetros de forma cerrada, se verifica el siguiente lema.

Lema 6.2 Sean $Y_n | X_n, \mathbf{b} \sim N_n(Z_n \mathbf{b}, \mathbf{s}^2 I_n)$ con $Z_n = (1_n : X_n)$ y $\mathbf{b} | \mathbf{s} \sim N_{k+1}(\mathbf{m}, \mathbf{s}^2 M^{-1})$ con $\mathbf{m}' = (\mathbf{m}_0, \dots, \mathbf{m}_k)$, $\mathbf{m}'_I = (\mathbf{m}_1, \dots, \mathbf{m}_k)$, $\mathbf{b}' = (\mathbf{b}_0, \dots, \mathbf{b}_k)$, $\mathbf{b}'_I = (\mathbf{b}_1, \dots, \mathbf{b}_k)$ y $M = \{m_{ij}\}$ para $i, j = 0, 1, \dots, k$ tales que

$$M = \begin{pmatrix} m_{00} & M'_{0I} \\ M_{0I} & M_{II} \end{pmatrix}$$

donde M_{II} es una matriz de $k \times k$. Entonces

$$x_{n+1}^{hb}(H) = \frac{1}{m_{00} + n} (M_{0I} + X'_n 1_n) + L_n(H) \left\{ M_{0I} \mathbf{m}_0 + M_{II} \mathbf{m}_I + X'_n Y_n - \frac{m_{00} \mathbf{m}_0 + M'_{0I} \mathbf{m}_I + 1'_n Y_n}{m_{00} + n} (M_{0I} + X'_n 1_n) \right\},$$

siendo

$$L_n(H) = \frac{E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] \text{COV}_n[\mathbf{b}_0, \mathbf{b}_I] + T - E_n[\mathbf{b}_0]}{\mathbf{s}^2 (E_n[\mathbf{b}_I]' V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I] + H)}.$$

Este lema nos ayudará a demostrar que en el caso del modelo lineal conjugado, el conjunto de puntos $x_{n+1}^{hb}(H)' = (x_{n1}(H), \dots, x_{nk}(H))$, obtenidos al variar H , está siempre sobre una línea recta que no depende ni de X_n ni de Y_n , y que por lo tanto, es la misma línea recta para cualquier estadio n . Este sorprendente resultado implica, que para distribuciones a prioris conjugadas, los puntos del diseño *certainty equivalence* y *miope* en cualquier estadio n , estarán siempre sobre una línea recta que depende únicamente de los parámetros de $\mathbf{p}_0(\mathbf{b})$. En la siguiente proposición se muestra la ecuación de dicha recta en función de los parámetros de $\mathbf{p}_0(\mathbf{b})$. A partir de ella se puede calcular $x_{n+1}^{hb}(H)$ en función de su componente j -ésima,

$x_{n+1,j}(H)$, lo que facilitará y agilizará la implementación de tal diseño, puesto que se convierte en un problema con una única variable de control.

Proposición 6.2 Sean $Y_n | X_n, \mathbf{b} \sim N_n(Z_n \mathbf{b}, \mathbf{s}^2 I_n)$ y $\mathbf{b} | \mathbf{s} \sim N_{k+1}(\mathbf{m}, \mathbf{s}^2 M^{-1})$ con $\mathbf{m}' = (\mathbf{m}_0, \mathbf{m}_1, \dots, \mathbf{m}_k)$, $\mathbf{m}'_l = (\mathbf{m}_1, \dots, \mathbf{m}_k)$ y sean $M = \{m_{ij}\}$, $i, j = 0, \dots, k$ particionada como en el Lema 6.2 y $\mathbf{m}'_j = (m_{j1}, \dots, m_{jk})$. Para cada $j = 1, \dots, k$ tal que $(m_{00} \mathbf{m}'_j - m_{j0} M'_{0l}) \mathbf{m}_l \neq 0$

$$x_{n+1}^{hb}(H) = (x_{n+1,1}(H), \dots, x_{n+1,k}(H))' = \frac{1}{m_{00}} M_{0l} + (x_{n+1,j}(H) - \frac{m_{j0}}{m_{00}}) e_j$$

con

$$e_j = \frac{1}{(m_{00} \mathbf{m}'_j - m_{j0} M'_{0l}) \mathbf{m}_l} (m_{00} M_{ll} - M_{0l} M'_{0l}) \mathbf{m}_l,$$

y

$$x_{n+1,j}(H) = \frac{m_{j0} + 1'_n x_j^n}{m_{00} + n} + L_n(H) \left\{ m_{j0} \mathbf{m}_0 + \mathbf{m}'_j \mathbf{m}_l + Y'_n x_j^n - \frac{m_{00} \mathbf{m}_0 + M'_{0l} \mathbf{m}_l + 1'_n Y_n (m_{j0} + 1'_n x_j^n)}{m_{00} + n} \right\}.$$

Cuando $(m_{00} M_{ll} - M_{0l} M'_{0l}) \mathbf{m}_l = \mathbf{s}^2 m_{00} V_0^{-1}[\mathbf{b}_l] \mathbf{m}_l = 0$ entonces $x_{n+1}^{hb}(H) = (1/m_{00}) M_{0l}$.

La demostración de esta proposición se encuentra en el Apéndice 6.A.3. Utilizando los resultados de inversión de matrices dados en Lema 2.1, se pueden expresar las componentes de la matriz M en función de la matriz de varianza-covarianza a priori como

$$m_{00} = - \frac{\mathbf{s}^2}{V_0[\mathbf{b}_0] - COV_0[\mathbf{b}_0, \mathbf{b}_l]' V_0^{-1}[\mathbf{b}_l] COV_0[\mathbf{b}_0, \mathbf{b}_l]},$$

$$M'_{0l} = m_{00} COV_0[\mathbf{b}_0, \mathbf{b}_l]' V_0^{-1}[\mathbf{b}_l],$$

$$M_{ll} = \mathbf{s}^2 V_0^{-1}[\mathbf{b}_l] + m_{00} V_0^{-1}[\mathbf{b}_l] COV_0[\mathbf{b}_0, \mathbf{b}_l] COV_0[\mathbf{b}_0, \mathbf{b}_l]' V_0^{-1}[\mathbf{b}_l].$$

En el supuesto de que los parámetros \mathbf{b} sean independientes entre sí, el resultado de la proposición anterior se puede simplificar en el siguiente corolario.

Corolario 6.2 Sean $Y_n | X_n, \mathbf{b} \sim N_n(Z_n \mathbf{b}, \mathbf{s}^2 I_n)$ y $\mathbf{b} | \mathbf{s} \sim N_{k+1}(\mathbf{m}, \mathbf{s}^2 M^{-1})$, con $V_0[\mathbf{b}_i] = \mathbf{s}^2_{b_i}$ y donde M es una matriz diagonal y así $COV_0[\mathbf{b}_i, \mathbf{b}_j] = 0$ para todo $i \neq j$. Si $\mathbf{m}_j \neq 0$ entonces

$$x_{n+1}^{hb}(H)' = (x_{n+1,1}(H), \dots, x_{n+1,k}(H)) = x_{n+1,j}(H) \frac{\mathbf{s}_{b_j}^2}{\mathbf{m}_j} \left(\frac{\mathbf{m}_1}{\mathbf{s}_{b_1}^2}, \dots, \frac{\mathbf{m}_k}{\mathbf{s}_{b_k}^2} \right).$$

Su demostración es inmediata haciendo uso de la Proposición 6.2 y teniendo en cuenta que $M'_{0I} = (0, \dots, 0)$, $\mathbf{m}'_j = (0, \dots, \mathbf{s}_{b_j}^{-2}, \dots, 0)$ y M_{II} es una matriz diagonal $k \times k$ de elementos $\mathbf{s}_{b_i}^{-2}$, para $i = 1, \dots, k$. Este resultado además indica que en este caso particular, la recta donde se encuentran todos los puntos *híbridos* es una recta que pasa por el origen, que además coincide con la recta que une *miope* con *certainty equivalence*.

6.2.3 Modelo lineal con priori conjugada (\mathbf{s}^2 desconocida).

Sean $Y_n | X_n, \mathbf{b} \sim N_n(Z_n \mathbf{b}, \mathbf{s}^2 I_n)$ pero donde \mathbf{s}^2 es desconocida y de la que se sabe que $1/\mathbf{s}^2 \sim Ga(\mathbf{a}, \mathbf{t})$ (o que \mathbf{s}^2 sigue una gamma inversa). Utilizando la formulación dada en Bernardo y Smith (1994), el modelo lineal conjugado asume que la distribución conjunta de los parámetros, $\mathbf{p}_0(\mathbf{b}, \mathbf{s}^2)$, sigue una normal-gamma, tal que

$$\mathbf{p}_0(\mathbf{b}, \mathbf{s}^2) = Ng(\mathbf{b}, \mathbf{s}^2 | \mathbf{m}, M, \mathbf{a}, \mathbf{t}) = N_{k+1}(\mathbf{b} | \mathbf{m}, \mathbf{s}^2 M^{-1}) Ga(1/\mathbf{s}^2 | \mathbf{a}, \mathbf{t}),$$

lo que hace que la marginal de \mathbf{b} sea una t-student multivariante:

$$\mathbf{p}_0(\mathbf{b}) = St_{k+1}(\mathbf{b} | \mathbf{m}, M \mathbf{a} \mathbf{t}^{-1}, 2\mathbf{a})$$

con $E_0[\mathbf{b}] = \mathbf{m}$ y $V_0[\mathbf{b}] = \mathbf{t}(\mathbf{a} - 1)^{-1} M^{-1}$. En estas circunstancias se tiene que la distribución marginal a posteriori de \mathbf{b} es nuevamente una t-student multivariante:

$$\mathbf{p}_n(\mathbf{b}) = St_{k+1}(\mathbf{b} | \mathbf{b}_n, (M + Z_n' Z_n)(\mathbf{a} + \frac{1}{2} n) \mathbf{t}_n^{-1}, 2\mathbf{a} + n)$$

con $\mathbf{b}_n = (M + Z_n' Z_n)^{-1} (M \mathbf{m} + Z_n' Y_n)$ y $\mathbf{t}_n = \mathbf{t} + ((Y_n - Z_n \mathbf{b}_n)' Y_n + (\mathbf{m} - \mathbf{b}_n)' M \mathbf{m}) / 2$, y cuyos momentos son: $E_n[\mathbf{b}] = \mathbf{b}_n$ y $V_n[\mathbf{b}] = \mathbf{t}_n (\mathbf{a} - 1 + n/2)^{-1} (M + Z_n' Z_n)^{-1}$. Comparando dichos momentos con los obtenidos cuando \mathbf{s}^2 se considera conocida, sigue que los resultados dados en el Lema 6.2, Proposición 6.2 y Corolario 6.2 son igualmente válidos con la única

modificación de que la \mathbf{s}^2 que presenta la expresión del $L_n(H)$ dada en la sección anterior debe ser sustituida por $\mathbf{t}_n(\mathbf{a} - 1 + n/2)^{-1}$.

6.2.4 Modelo lineal de segundo orden.

En un contexto más general, se podría considerar un modelo lineal de segundo orden con interacciones dobles, con k controles, x_1, \dots, x_k , y $l+1$ parámetros $\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_k, \mathbf{b}_{12}, \dots, \mathbf{b}_{k-1,k}$, de la forma

$$E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{j=1}^k \mathbf{b}_j x_j + \sum_{\substack{j, j' \in K \\ j < j'}} \mathbf{b}_{j, j'} x_j x_{j'},$$

con $K = \{1, \dots, k\}$ y no necesariamente con todas las interacciones. En la práctica, en la mayoría de los problemas no será necesario suponer un modelo tan complejo pero de todas maneras indicamos a continuación cómo se calcularía $x_{n+1}^{hb}(H)$ en los mismos. Para este tipo de modelo la expresión del $x_{n+1}^{hb}(H)$ no se obtiene de forma cerrada y su valor pasa por la resolución del sistema de ecuaciones no lineales que en la siguiente proposición se introduce.

Proposición 6.3 *Sea el modelo con*

$$E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{j=1}^k \mathbf{b}_j x_j + \sum_{\substack{j, j' \in K \\ j < j'}} \mathbf{b}_{j, j'} x_j x_{j'},$$

y con $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$, $K = \{1, \dots, k\}$, $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}'_1) = (\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_k, \mathbf{b}_{12}, \dots, \mathbf{b}_{k-1,k}) \sim \mathbf{p}_0(\mathbf{b})$, independiente de \mathbf{e} . Sean además A y b , una matriz cuadrada $l \times l$ ($l+1$ número de parámetros) y un vector $l \times 1$, respectivamente, dados por

$$A = E_n[\mathbf{b}_l] E_n'[\mathbf{b}_l] + H V_n[\mathbf{b}_l]$$

y

$$b = (E_n[\mathbf{b}_0] - T) E_n[\mathbf{b}_l] + H COV_n[\mathbf{b}_0, \mathbf{b}_l].$$

Consideremos la matriz A particionada como

$$A = \begin{pmatrix} A_{11} & A'_{12} \\ A_{12} & A_{22} \end{pmatrix}$$

tal que $A_{11} = \{a_{j,t}\}$, $A_{12} = \{a_{t,jj'}\}$ y $A_{22} = \{a_{jj',tt'}\}$ con $j, j', t, t' \in K$, $j < j'$ y $t < t'$, submatrices de $k \times k$, $(l-k) \times k$ y de $(l-k) \times (l-k)$ respectivamente. Entonces el diseño $d_{hb}(H)$ observa y_{n+1} en el $x_{n+1}^{hb}(H)$ que resuelve

$$\begin{aligned} \min \{hb_{n+1}(x; H)\} = \min \{ & \sum_{j \in K} (\sum_{t \in K} a_{j,t} x_t) x_j + 2 \sum_{\substack{j \in K \\ t, t' \in K \\ t < t'}} (\sum_{\substack{j, j' \in K \\ j < j'}} a_{j,tt'} x_t x_{t'}) x_j + \sum_{\substack{j, j' \in K \\ j < j'}} (\sum_{\substack{t, t' \in K \\ t < t'}} a_{tt',jj'} x_t x_{t'}) x_j x_{j'} \\ & + 2 \sum_{j \in K} b_j x_j + \sum_{\substack{j, j' \in K \\ j < j'}} b_{j,j'} x_j x_{j'} \}, \end{aligned}$$

o, de forma equivalente, el sistema

$$\begin{aligned} F_i(x_1, \dots, x_k) = & a_{i,i} x_i + \sum_{t \in K - \{i\}} (a_{i,t} + 2 a_{i,it} x_i) x_t + \sum_{j, t \in K - \{i\}} (a_{j,it} + a_{ij,it} x_i) x_t x_j + \\ & \sum_{\substack{t, t' \in K - \{i\} \\ t < t'}} (a_{i,tt'} + \sum_{j \in K - \{i\}} a_{ij,tt'} x_j) x_t x_{t'} + b_i + \sum_{j \in K - \{i\}} b_{i,j} x_j = 0, \end{aligned}$$

para $i \in K$.

La resolución de dicho sistema no lineal se puede llevar a cabo utilizando el método de Newton-Raphson descrito en el Apéndice 6.A.4. La matriz jacobiana asociada a dicho método se encuentra en el Lema 6.6 del Apéndice 6.A.2 y una propuesta de punto de inicio para aplicar dicho método se anota al final de este mismo apéndice.

6.3 Aplicación al modelo lineal.

Para ilustrar el uso de los diseños *híbridos*, se explora su utilización en el modelo lineal normal con $f(x; \mathbf{b}) = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i$. En esta sección se representa para una priori concreta la evolución de la pérdida esperada en función de H , tanto para un modelo lineal simple, con una sola variable de control, como para un modelo con tres variables de control.

6.3.1 Ejemplo para modelo lineal simple ($k = 1$).

La Figura 6.1 muestra la evolución de $r_2(H)$ para el diseño *híbrido* en función de H cuando la priori es tal que $\mathbf{b}_0 \sim N(0., \mathbf{s}_{b_0}^2 = .5^2)$ y $\mathbf{b}_1 \sim N(1., \mathbf{s}_{b_1}^2 = 1.^2)$ con \mathbf{b}_0 independiente de \mathbf{b}_1 , $\mathbf{s} = .5$ y $N = 20$. El gráfico de la izquierda corresponde a $R = 1$ y el de la derecha a $R = 10$. Los valores de $r_2(H = 0)$ y $r_2(H = 1)$, en los extremos de la gráfica, corresponden a las pérdidas esperadas del diseño *certainty equivalence* y *miope*, respectivamente.

En dichos gráficos primero se observa que el valor de H^* depende muy poco del tamaño de la región experimental, como se volverá a apreciar de forma más general en la Sección 6.4.

También queda patente que el valor de $r_2(H = 0)$ empeora sustancialmente al aumentar la región experimental, mientras que el valor de $r_2(H = 1)$ y el de $r_2(H^*)$ mejoran ligeramente.

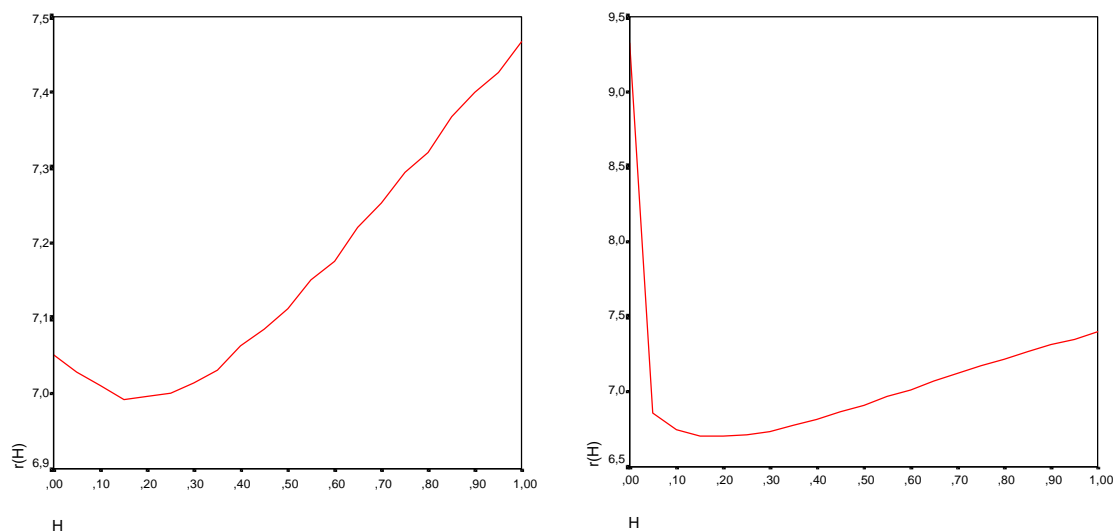


Figura 6.1. Comportamiento de $r(H)$ del diseño híbrido cuando $(y|x, \mathbf{b}) \sim N(\mathbf{b}_0 + \mathbf{b}_1 x, \mathbf{s}^2)$, con $\mathbf{b}_0 \sim N(0., \mathbf{s}_{b_0}^2 = .5^2)$, $\mathbf{b}_1 \sim N(1., \mathbf{s}_{b_1}^2 = 1.^2)$, \mathbf{b}_1 independiente de \mathbf{b}_0 , $T = 0$ y $\mathbf{s} = .5$. El gráfico de la izquierda para la región experimental $C = [-1., 1.]$ y el de la derecha para $C = [-10., 10.]$

Esto indica que la eficiencia de *certainty equivalence* degenera mucho al aumentar el tamaño de la región, seguramente debido a la gran variabilidad del valor $(T - E_n[\mathbf{b}_0])/E_n[\mathbf{b}_1]$, que lleva a dicho diseño a experimentar con probabilidad no despreciable en las fronteras de la región. Todo esto volverá a quedar reflejado con más generalidad en la Sección 6.4.

6.3.2 Ejemplo para modelo lineal múltiple ($k = 3$).

Consideremos ahora un ejemplo con tres controles ($k = 3$). La Figura 6.2 muestra la evolución de $r_2(H)$ en función de H para la priori $\mathbf{b}_0 \sim N(0., \mathbf{s}_{\mathbf{b}_0}^2 = .5^2)$ y $\mathbf{b}_i \sim N(1., \mathbf{s}_{\mathbf{b}_i}^2 = 1.^2)$, $i = 1, \dots, 3$, \mathbf{b}_0 independiente de \mathbf{b}_i , $\mathbf{s} = .5$, $N = 20$. El gráfico de la izquierda corresponde a $R = 1$ y el gráfico de la derecha a $R = 10$. Tal y como ya observamos en la Figura 6.1, los valores de $r(H = 0)$ y $r(H = 1)$ corresponden a los puntos en los extremos de la gráfica. Como se aprecia, su comportamiento es bastante similar al caso de un único control.

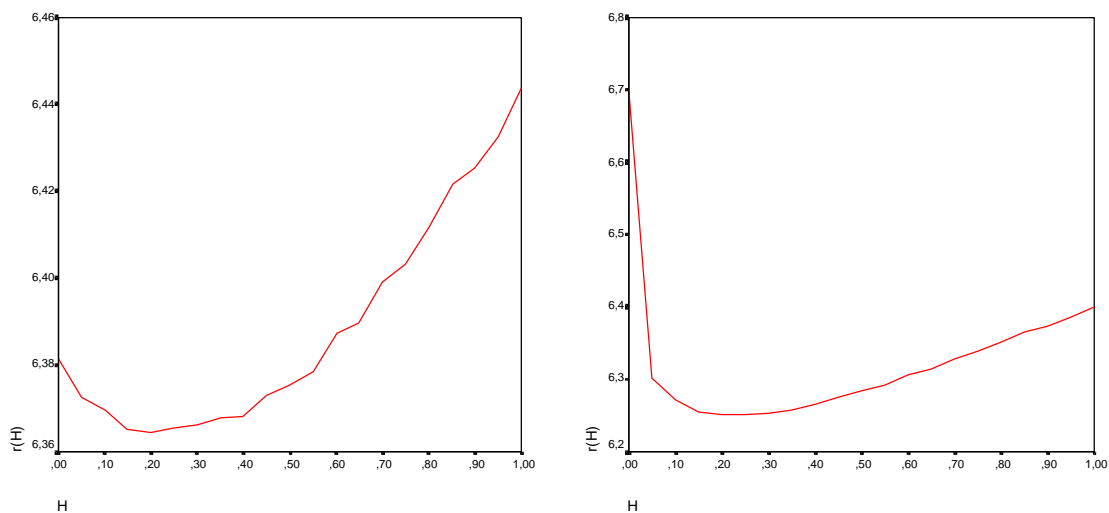


Figura 6.2. Comportamiento de $r(H)$ del diseño híbrido cuando $(y|x, \mathbf{b}) \sim N(\mathbf{b}_0 + \sum_{i=1}^3 \mathbf{b}_i x_i, \mathbf{s}^2)$, con $\mathbf{b}_0 \sim N(0., \mathbf{s}_{\mathbf{b}_0}^2 = .5^2)$, $\mathbf{b}_i \sim N(1., \mathbf{s}_{\mathbf{b}_i}^2 = 1.^2)$, $i = 1, 2, 3$, \mathbf{b}_i independientes entre ellos y de \mathbf{b}_0 , $T = 0$ y $\mathbf{s} = .5$. El gráfico de la izquierda para la bola de radio 1. y el de la derecha de radio 10.

6.4 Análisis de sensibilidad.

En la Sección 6.3 hemos presentado el comportamiento de $r_2(H)$ en función de H para una priori concreta. Aquí en cambio estudiaremos el comportamiento de H^* , así como la mejora relativa de $d_{hb}(H^*)$, en función de la priori, N y tamaño de C , todo para pérdida cuadrática. Como ejemplo del comportamiento de los diseños *híbridos* se usará el modelo lineal normal con priori $\mathbf{p}_0(\mathbf{b})$ tal que $\mathbf{b}_0 \sim N(0., \mathbf{S}_{b_0}^2 = .5^2)$, $\mathbf{b}_i \sim N(1., \mathbf{S}_{b_i}^2)$ para $i=1, \dots, k$, y donde las \mathbf{b}_j son independientes para $j=0, 1, \dots, k$. La región experimental C es la bola de radio R centrada en el origen y, sin pérdida de generalidad, se considerará $T=0$. Cuando al computar $x_{n+1}^{hb}(H)$ éste caiga fuera de la región C se experimentará en el punto x_{n+1} de la frontera de C que minimiza $hb_{n+1}(H)$, tal como se detalla en el Apéndice 6.A.4.

Para computar $r_2(H)$ se simulará 1.000.000 de realizaciones de $\mathbf{b}' = (\mathbf{b}_0, \dots, \mathbf{b}_k)$ siguiendo $\mathbf{p}_0(\mathbf{b})$, y para cada \mathbf{b} se ejecuta el diseño $d_{hb}(H)$ sobre las $y_i(x_i^{hb})$ simuladas de la distribución de $(y_i | x_i^{hb}, \mathbf{b})$, promediándose las pérdidas observadas. La desviación estándar estimada para $\hat{r}_2(H)$ es del orden de .002 y el $H = H^*$ minimizando $r_2(H)$ puede ser computado fácilmente utilizando el método de Brent (1973), implementado en Press et al (1992).

Para evaluar la mejora relativa del mejor diseño *híbrido*, $d_{hb}(H^*)$, con respecto a cualquier diseño *híbrido*, $d_{hb}(H)$, definimos $rr(H)$ como la diferencia relativa entre la pérdida esperada bajo $d_{hb}(H^*)$ y bajo $d_{hb}(H)$,

$$rr(H) = \frac{r_2(H) - r_2(H^*)}{r_2(H^*)}.$$

Notar que $rr(0)$ y $rr(1)$ son las mejoras relativas de $d_{hb}(H^*)$ sobre *certainty equivalence* y *miope*, respectivamente. A continuación, se ejemplifican distintas situaciones para el caso de un único control ($k=1$) y para tres variables de control ($k=3$).

6.4.1 Ejemplo para modelo lineal simple ($k = 1$).

La Figura 6.3 explora como evolucionan H^* , $rr(0)$ y $rr(1)$ dependiendo de los valores de \mathbf{s}_{b_1} , de \mathbf{s} , del horizonte N y del tamaño de la región experimental R . Como se aprecia en dicha figura, cuando $N = 20$ y $R = 10$, H^* aumenta cuando aumenta \mathbf{s} y disminuye \mathbf{s}_{b_1} , aunque es algo menos sensible a las variaciones de \mathbf{s}_{b_1} . Esto se traduce en que cuanto menos conozcamos los parámetros \mathbf{b} y mayor es el ruido de $y|x$, más nos acercamos al diseño *miope*. También sigue que la elección del mejor *híbrido* depende mucho más de \mathbf{s} que de \mathbf{s}_{b_1} .

Además, cuando N crece, H^* disminuye alejándose de 1 hacia 0. Esto es consistente con el hecho de que por un lado los diseños *miopes* se deterioran cuando N crece y por otro con que los diseños *certainty equivalence* Bayes son asintóticamente optimales tal como demuestran Chen y Hu (1998). Es importante notar que H^* es insensible a los cambios en R y, por tanto, la elección del mejor diseño *híbrido* está muy poco influenciada por el tamaño de la región experimental.

La Figura 6.3 también muestra como tanto $rr(0)$ como $rr(1)$ aumentan cuando aumenta \mathbf{s}_{b_1} o disminuye \mathbf{s} ; Cuando menos se conoce sobre \mathbf{b}_1 y menor es la variabilidad de y sobre su media, mayor será la mejora de $d_{hb}(H^*)$ sobre *certainty equivalence* y *miope*. Más aún se observa que la mejora relativa respecto a *certainty equivalence* acostumbra a ser notablemente mayor que para *miope*, lo que indirectamente indica que *miope* acostumbra a ser más eficiente que *certainty equivalence*.

Por otro lado, cuando N aumenta, las mejoras relativas de $d_{hb}(H^*)$ sobre *miope*, $rr(1)$, aumentan, pero sobre *certainty equivalence*, $rr(0)$, disminuye como se esperaba. Las mejoras sobre *miope* son mucho menos sensibles a las variaciones de R que las mejoras sobre

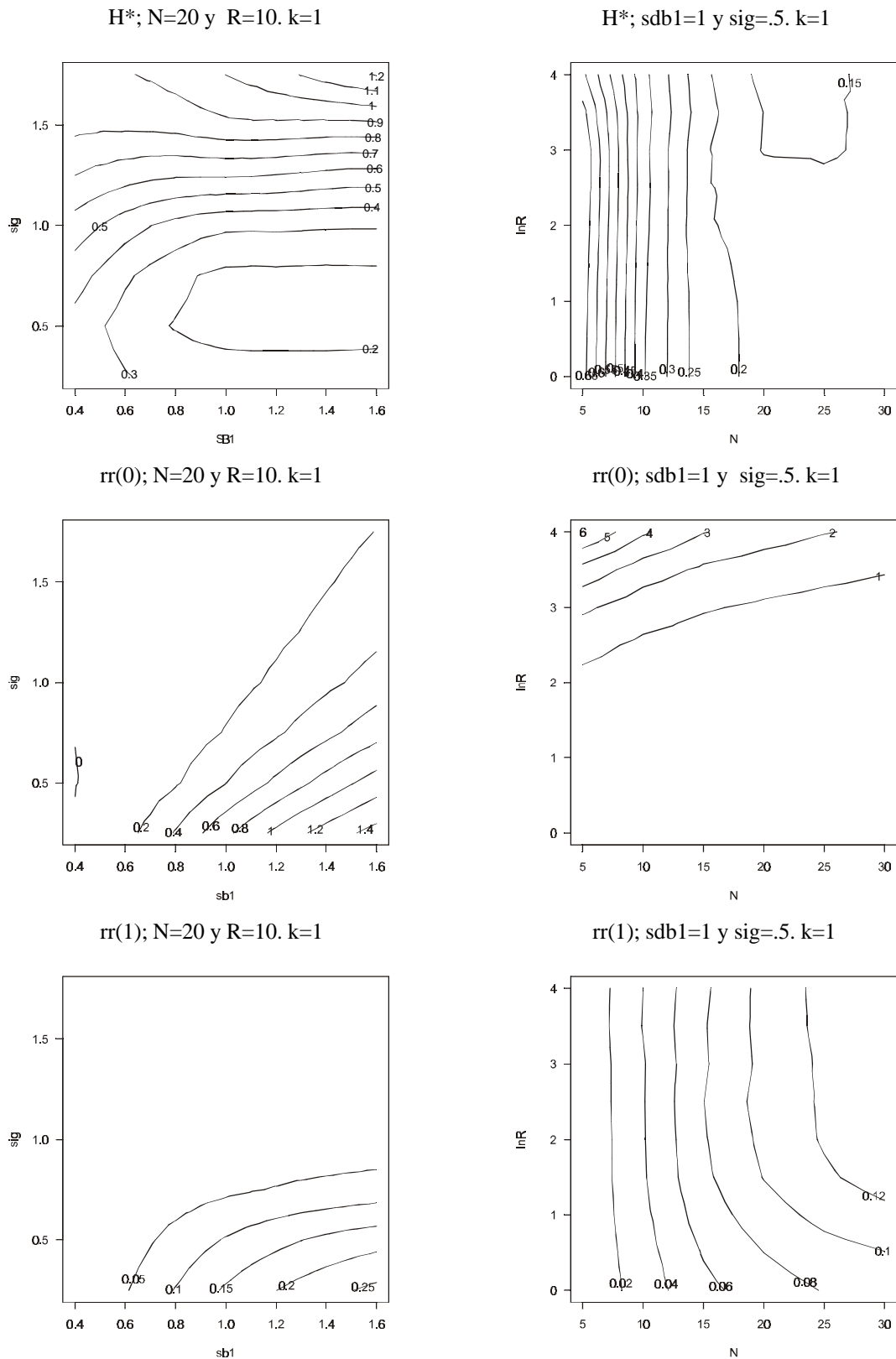


Figura 6.3. Superficies de contorno de H^* y de la mejora relativa de $d_{hb}(H^*)$ respecto a c.e. y miope, $rr(0)$ y $rr(1)$, cuando $(y|x, \mathbf{b}) \sim N(\mathbf{b}_0 + \mathbf{b}_1 x, \mathbf{s}^2)$, con $\mathbf{b}_0 \sim N(0., \mathbf{s}_{b_0}^2 = .5^2)$, $\mathbf{b}_1 \sim N(1., \mathbf{s}_{b_1}^2)$, \mathbf{b}_1 independiente de \mathbf{b}_0 . La región experimental es $C = [-R, R]$, $T = 0$ y $sb1$ y sig representando $\mathbf{s}_{b_1}^2$ y \mathbf{s} , respectivamente.

certainty equivalence debido a la fuerte dependencia de *certainty equivalence* sobre R . En particular, *certainty equivalence* tiene un comportamiento extremadamente pobre para regiones experimentales grandes.

6.4.2 Ejemplo para modelo lineal múltiple con $k = 3$.

La Figura 6.4 muestra como evoluciona H^* , $rr(0)$ y $rr(1)$ dependiendo de los valores de \mathbf{s}_{b_1} , de \mathbf{s} , del horizonte N y del tamaño de la región experimental R , de forma similar a la Figura 6.3, pero en este caso para tres variables control. Las conclusiones y lecturas de los gráficos son también prácticamente idénticas a las dadas para la Figura 6.3, es decir, que H^* , $rr(0)$ y $rr(1)$ tienen un comportamiento similar a través de diferentes k 's, siendo las mejoras relativas de $d_{hb}(H^*)$ sobre *certainty equivalence* y *miopie* menores cuanto más grande es k .

La mejora relativa respecto a *certainty equivalence* es menos pronunciada para $k = 3$ que para $k = 1$, en parte debido a que para conseguir unicidad en la elección *certainty equivalence* hemos introducido la miopiedad. Para los mismos valores de N y R ó \mathbf{s} y \mathbf{s}_{b_1} en las Figuras 6.3 y 6.4, se observa que los valores de H^* en esta última son ligeramente superiores a los de la Figura 6.3, lo que sugiere mayor proximidad hacia la elección *miopie* y quizás que la elección *certainty equivalence* en el caso de un mayor número de variables de control será una buena opción sólo cuando N es suficientemente grande.

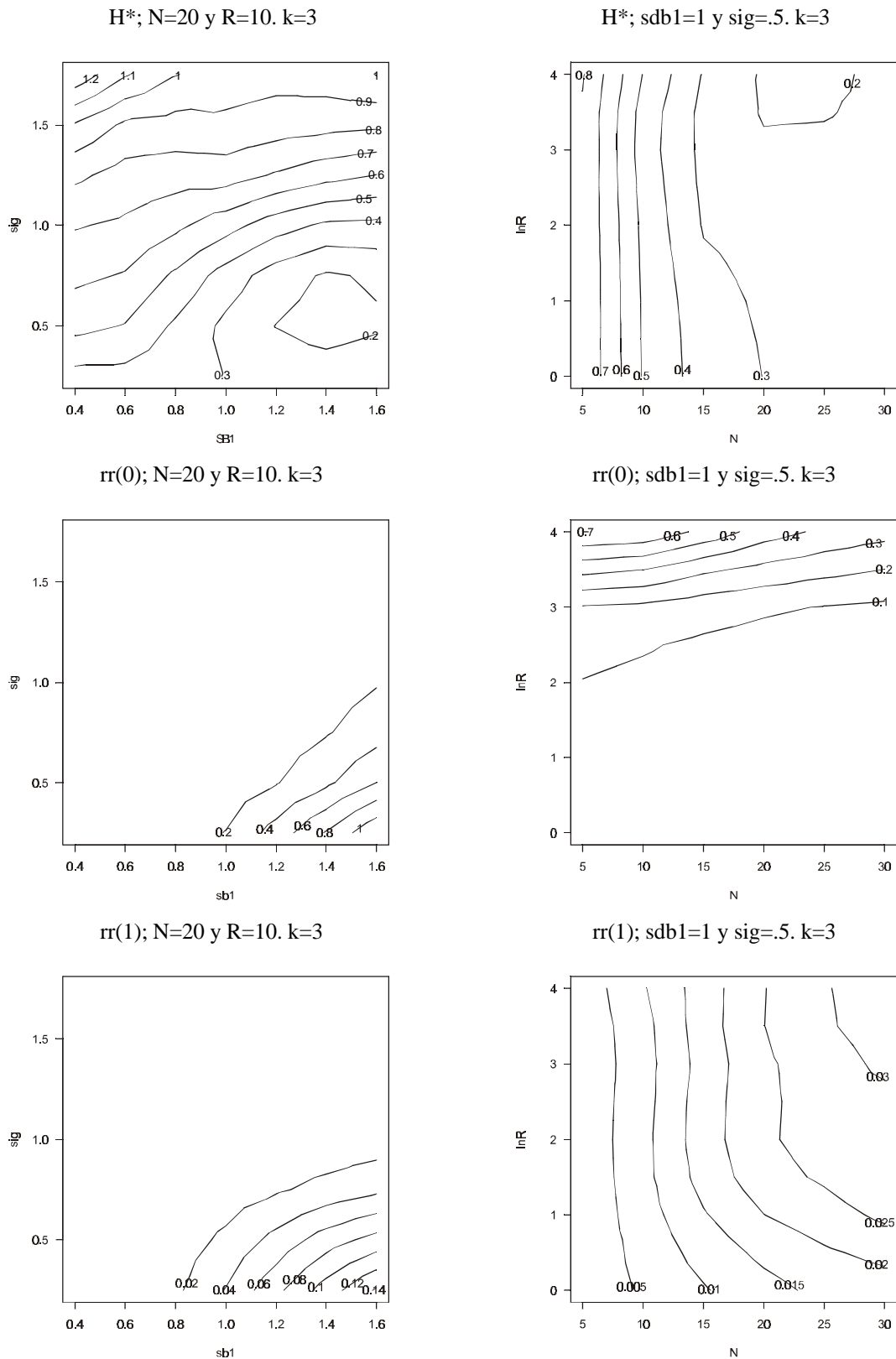


Figura 6.4. Superficies de contorno de H^* y de la mejora relativa de $d_{hb}(H^*)$ respecto a c.e. y miope, $rr(0)$ y $rr(1)$, cuando $(y|x, \mathbf{b}) \sim N(\mathbf{b}_0 + \sum_{i=1}^3 \mathbf{b}_i x_i, \mathbf{S}^2)$, con $\mathbf{b}_0 \sim N(0, \mathbf{S}_{b_0}^2 = .5^2)$, $\mathbf{b}_i \sim N(1, \mathbf{S}_{b_i}^2)$, $i=1,2,3$, y con \mathbf{b}_i independiente entre ellas y de \mathbf{b}_0 . La región experimental C es una bola de radio R centrada en el origen y $T=0$.

6.5 Extensiones y comentarios.

Los diseños *híbridos* usan el paradigma bayesiano para inducir una acción de exploración sobre el diseño *miope* que ayuda a mejorar los estimadores de \mathbf{b} de los estadios intermedios y así conseguir un mejor control futuro. Computando H^* se encuentra un buen compromiso entre la elección *certainty equivalence* y la elección *miope*. Para obtener mejoras adicionales partiendo del mejor diseño *híbrido* siempre se puede recomputar H^* en algunos o en todos los estadios intermedios m , H_m^* , usando la distribución a posteriori de \mathbf{b} en ese estadio. Para el modelo lineal simple implementamos dicha alternativa actualizando en todos los estadios así como en sólo alguno de ellos, pero las mejoras obtenidas sobre el diseño *híbrido* eran mínimas, siendo el coste computacional bastante más elevado. Creemos que esa podría ser una buena estrategia para problemas con más de una variable de control, aunque para verificarlo es preciso contar con máquinas de computo más potentes.

Tal y como se ha descrito, cuando se selecciona el mejor diseño *híbrido* a través de simulación se usa el mismo modelo y la misma priori para construir la familia de diseños *híbridos*, pero en algunos contextos se podrían hacer de maneras alternativas. En particular, se podría usar el modelo lineal para generar el conjunto de diseños *híbridos*, computando así $x_{n+1}^{hb}(H)$ a través de la Proposición 6.1, pero en cambio computar H^* simulando las respuestas a partir de una familia de modelos más complicada e hipotéticamente más “realista”. Por ejemplo, esto podría ser lo más apropiado cuando se trabaja con datos provenientes de respuesta cuantales o modelables a través de modelos lineales con interacciones y términos cuadráticos o modelos lineales generalizados.

Al construir la familia de diseños *híbridos* a través de la minimización de una combinación convexa de los criterios de optimización que definen los diseños *certainty equivalence* y *miope*, se incluyen en esa familia estos dos diseños. De forma análoga se podrían

construir familias que incluyeran cualquier grupo de r diseños dados. La nueva familia estaría indexada por un vector de índices $\bar{H} = (H_1, H_2, \dots, H_{r-1})$ y computando la mejor $\bar{H}^* = (H_1^*, H_2^*, \dots, H_{r-1}^*)$ se obtendrían diseños que mejoran cada uno de los r diseños usados para construir la familia. Esta técnica puede ser implementada en cualquier problema de diseño óptimo, tanto en problemas secuenciales como no secuenciales, donde la estrategia optimal es demasiado difícil de computar.

Apéndice 6.A.1 Varianzas de predicción.

El siguiente lema muestra la relación existente entre las varianzas de predicción en los puntos seleccionados por los diseños *híbridos*, *certainty equivalence* y *miope*.

Lema 6.3 Sean $y_m = y(x_{n+1}^{my})$, $y_c = y(x_{n+1}^{ce})$ e $y_h^H = y(x_{n+1}^{hb}(H))$. Cuando $0 < H < 1$,

$$V_n[y_m] < V_n[y_c] + \frac{E_n[(y_h^H - T)^2] - E_n[(y_m - T)^2]}{1 - H} < V_n[y_h^H],$$

$$V_n[y_h^H] < V_n[y_c] + \frac{(E_n[y_h^H - T])^2 - (E_n[y_c - T])^2}{H} < V_n[y_c].$$

En particular,

$$V_n[y_m] < V_n[y_h^H] < V_n[y_c].$$

Demostración.

Por definición de los diseños *certainty equivalence* y *miope* se verifica que, para cualquier $H \in [0, 1]$,

$$E_n[(y_m - T)^2] \leq E_n[(y_h^H - T)^2],$$

$$0 = (E_n[y_c - T])^2 \leq (E_n[y_h^H - T])^2.$$

Además, por la definición de diseños *híbridos*, se tiene que

$$E_n[(y_h^H - T)^2] + (H - 1)V_n[y_h^H - T] \leq E_n[(y_m - T)^2] + (H - 1)V_n[y_m - T],$$

$$(E_n[y_h^H - T])^2 + H V_n[y_h^H - T] \leq (E_n[y_c - T])^2 + H V_n[y_c - T],$$

o lo que es equivalente

$$E_n[(y_h^H - T)^2] - E_n[(y_m - T)^2] \leq (H - 1)(V_n[y_m - T] - V_n[y_h^H - T]),$$

$$(E_n[y_h^H - T])^2 - (E_n[y_c - T])^2 \leq H(V_n[y_c - T] - V_n[y_h^H - T]).$$

Ahora el resultado del lema se obtiene cuando $0 < H < 1$. ■

Apéndice 6.A.2 Demostración Proposición 6.1 y 6.3.

El siguiente lema presenta $hb_{n+1}(x; H)$ para modelos lineales en general con un formato útil para poder derivar con respecto a x , que sirve incluso para modelos con interacciones y con términos cuadráticos.

Lema 6.4 Sea $E[y|x, \mathbf{b}] = z'\mathbf{b} = \mathbf{b}_0 + x'\mathbf{b}_I$ donde $x' = (x_1, \dots, x_I)$, $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$ $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_I) = (\mathbf{b}_0, \mathbf{b}'_I)$ independiente de \mathbf{e} y $(\mathbf{b} : \mathbf{s}^2) \sim \mathbf{p}_0(\mathbf{b}, \mathbf{s}^2)$. Entonces

$$hb_{n+1}(x; H) = x'(E_n[\mathbf{b}_I]E_n[\mathbf{b}_I]' + H V_n[\mathbf{b}_I])x +$$

$$2x'((E_n[\mathbf{b}_0] - T)E_n[\mathbf{b}_I] + H COV_n[\mathbf{b}_0, \mathbf{b}_I]) + E_n^2[\mathbf{b}_0] - 2T E_n[\mathbf{b}_0] + T^2 + H V_n[\mathbf{b}_0] + \hat{\mathbf{s}}_n^2 H.$$

Demostración.

Aplicando el Lema 6.1 con $f(x; \mathbf{b}) = z'\mathbf{b}$, donde $z' = (1|x') = (1, x_1, \dots, x_I)$, se tiene que

$$hb_{n+1}(x; H) = E_n^2[z'\mathbf{b}] - 2T E_n[z'\mathbf{b}] + T^2 + H V_n[z'\mathbf{b}] + \hat{\mathbf{s}}_n^2 H =$$

$$z'(E_n[\mathbf{b}]E_n[\mathbf{b}]' + H V_n[\mathbf{b}])z - 2T E_n[\mathbf{b}]'z + T^2 + \hat{\mathbf{s}}_n^2 H$$

y esto puede ser reescrito como está en el lema. ■

El siguiente lema presenta la ecuación de $x_{n+1}^{hb}(H)$ cuando $H \in (0, 1]$ para un modelo lineal de primer orden.

Lema 6.5 Sean $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \sum_{i=1}^k \mathbf{b}_i x_i$, $V[\mathbf{e}|x, \mathbf{b}] = \mathbf{s}^2$ y $\mathbf{b}' = (\mathbf{b}_0, \dots, \mathbf{b}_k)$ independiente de \mathbf{e} y $(\mathbf{b} : \mathbf{s}^2) \sim \mathbf{p}_0(\mathbf{b}, \mathbf{s}^2)$. Cuando $H \in (0, 1]$, el diseño $d_{hb}(H)$ observa y_{n+1} en

$$x_{n+1}^{hb}(H) = -(E_n[\mathbf{b}_I]E_n[\mathbf{b}_I]' + H V_n[\mathbf{b}_I])^{-1}((E_n[\mathbf{b}_0] - T)E_n[\mathbf{b}_I] + H COV_n[\mathbf{b}_0, \mathbf{b}_I]),$$

donde $\mathbf{b}'_I = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k)$ es el vector de todos los parámetros excepto \mathbf{b}_0 y $COV_n[\mathbf{b}_0, \mathbf{b}_I]$ es el vector $k \times 1$ de covarianzas a posteriori entre \mathbf{b}_0 y los k elementos de \mathbf{b}_I .

Demostración.

El resultado sigue de derivar $hb_{n+1}(x; H)$ dado en Lema 6.4 con respecto a x e igualarlo a cero, y resolver el sistema de ecuaciones lineales resultante. ■

Cuando $H = 0$ y $k > 1$, $(E_n[\mathbf{b}_I]E_n[\mathbf{b}_I]')$ no es de rango máximo y este lema no podría ser utilizado, pero $x_{n+1}^{hb}(H=0)$ se puede seguir calculando sin más que reemplazar la inversa por una inversa generalizada.

Demostración Proposición 6.1.

Aplicando la fórmula de Sherman-Morrison

$$(A + uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1 + v'A^{-1}u}$$

se tiene que

$$(H V_n[\mathbf{b}_I] + E_n[\mathbf{b}_I]E_n[\mathbf{b}_I]')^{-1} = \frac{V_n^{-1}[\mathbf{b}_I]}{H} - \frac{V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I]E_n[\mathbf{b}_I]V_n^{-1}[\mathbf{b}_I]}{H(H + E_n[\mathbf{b}_I]V_n^{-1}[\mathbf{b}_I]E_n[\mathbf{b}_I])}$$

Para $H \in (0, 1]$, utilizando esta última expresión y desarrollando la ecuación de $x_{n+1}^{hb}(H)$ dada en el Lema 6.5, se obtiene la fórmula deseada.

Cuando $H = 0$, al sustituir H por cero en la expresión del $x_{n+1}^{hb}(H)$ dada en esta proposición, coincide con la dada en la Proposición 2.1 al definir el *certainty equivalence* para un modelo lineal de este tipo. ■

Demostración Proposición 6.3.

Utilizando el Lema 6.4 y descomponiendo $x' = ((x^1)', (x^2)')$ con $(x^1)' = (x_1, \dots, x_k)$ y $(x^2)' = (x_1 x_2, \dots, x_{k-1} x_k)$, se tiene que

$$\min \{hb_{n+1}(x; H)\} = \min \{x'Ax + 2x'b\} = \min \{(x^1)'A_{11}x^1 + 2(x^2)'A_{12}x^1 + (x^2)'A_{22}x^2 + 2x^1b\},$$

de forma que

$$(x^1)'A_{11}x^1 = \sum_{j \in K} \left(\sum_{t \in K} a_{j,t} x_t \right) x_j = a_{i,i} x_i^2 + 2 \sum_{\substack{j,t \in K \\ j < t}} a_{j,t} x_t x_j = a_{i,i} x_i^2 + 2 x_i \sum_{t \in K-\{i\}} a_{i,t} x_t + 2 \sum_{\substack{j,t \in K-\{i\} \\ j < t}} a_{j,t} x_t x_j .$$

$$(x^2)'A_{12}x^1 = \sum_{j \in K} \left(\sum_{\substack{t,t' \in K \\ t < t'}} a_{j,tt'} x_t x_{t'} \right) x_j = x_i \sum_{\substack{t,t' \in K \\ t < t'}} a_{i,tt'} x_t x_{t'} + \sum_{\substack{j \in K-\{i\} \\ t,t' \in K \\ t < t'}} \left(\sum a_{j,tt'} x_t x_{t'} \right) x_j = \\ x_i^2 \sum_{t \in K-\{i\}} a_{i,ii} x_t + x_i \sum_{\substack{t,t' \in K-\{i\} \\ t < t'}} a_{i,tt'} x_t x_{t'} + x_i \sum_{j \in K-\{i\}} \left(\sum_{t \in K-\{i\}} a_{j,it} x_t \right) x_j + \sum_{j \in K-\{i\}} \left(\sum_{\substack{t,t' \in K-\{i\} \\ t < t'}} a_{j,tt'} x_t x_{t'} \right) x_j .$$

$$(x^2)'A_{22}x^2 = \sum_{\substack{j,j' \in K \\ j < j'}} \left(\sum_{\substack{t,t' \in K \\ t < t'}} a_{tt',jj'} x_t x_{t'} \right) x_j x_{j'} = x_i^2 \sum_{j \in K-\{i\}} \left(\sum_{t \in K-\{i\}} a_{ij,it} x_t \right) x_j + x_i \sum_{j \in K-\{i\}} \left(\sum_{\substack{t,t' \in K-\{i\} \\ t < t'}} a_{ij,tt'} x_t x_{t'} \right) x_j + \\ x_i \sum_{\substack{j,j' \in K-\{i\} \\ j < j'}} \left(\sum_{t \in K-\{i\}} a_{jj',ij} x_t \right) x_j x_{j'} + \sum_{\substack{j,j' \in K-\{i\} \\ j < j'}} \left(\sum_{\substack{t,t' \in K-\{i\} \\ t < t'}} a_{jj',tt'} x_t x_{t'} \right) x_j x_{j'} .$$

$$x'b = \sum_{j \in K} b_j x_j + \sum_{\substack{j,j' \in K \\ j < j'}} b_{j,j'} x_j x_{j'} = b_i x_i + \sum_{j \in K-\{i\}} b_j x_j + x_i \sum_{j \in K-\{i\}} b_{i,j} x_j + \sum_{\substack{j,j' \in K-\{i\} \\ j < j'}} b_{j,j'} x_j x_{j'} .$$

Derivando con respecto a x_i , para $i \in K = \{1, \dots, k\}$, e igualando a cero, se obtiene el sistema no lineal de la proposición. ■

En el supuesto de utilizar el método de Newton-Raphson, introducido en el Apéndice 6.A.4, para la resolución del sistema,

$$F_i(x_1, \dots, x_k) = a_{i,i} x_i + \sum_{t \in K-\{i\}} (a_{i,t} + 2 a_{i,ii} x_i) x_t + \sum_{j,t \in K-\{i\}} (a_{j,it} + a_{ij,it} x_i) x_t x_j + \\ \sum_{\substack{t,t' \in K-\{i\} \\ t < t'}} (a_{i,tt'} + \sum_{j \in K-\{i\}} a_{ij,tt'} x_j) x_t x_{t'} + b_i + \sum_{j \in K-\{i\}} b_{i,j} x_j = 0, \quad (6.1)$$

con $i \in K$, la matriz jacobiana asociada, $J = \{J_{ir}\}$ queda definida en el siguiente lema.

Lema 6.6 La matriz jacobiana $J = \{J_{ir}\}$ con $J_{ir} \equiv \partial F_i / \partial x_r$, para $i, r \in K = \{1, \dots, k\}$, asociada al sistema $F_i(x_1, \dots, x_k) = 0$ dado en (6.1) es:

$$J_{ii} = a_{i,i} + 2 \sum_{j \in K-\{i\}} a_{i,ij} x_j + \sum_{j,t \in K-\{i\}} a_{ij,it} x_j x_t ,$$

$$J_{ir} |_{i \neq r} = a_{i,r} + 2 a_{i,ir} x_i + 2 x_r (a_{r,ir} + a_{ir,ir} x_i + \sum_{j \in K - \{i,r\}} a_{ir,jr} x_j) + \sum_{j \in K - \{i,r\}} (a_{r,ij} + a_{j,ir} + a_{i,jr} + 2 a_{ir,ij} x_i + \sum_{t \in K - \{i,r\}} a_{it,jr} x_t) x_j + \sum_{\substack{t,t' \in K - \{i,r\} \\ t < t'}} a_{ir,tt'} x_t x_{t'} + b_{i,r}.$$

Demostración.

J_{ii} se obtiene derivando con respecto a x_i y J_{ir} , con $i \neq r$, desarrollando los sumandos de F_i como sigue y derivando con respecto a x_r :

$$\begin{aligned} \sum_{j,t \in K - \{i\}} (a_{j,it} + a_{ij,it} x_i) x_t x_j &= \sum_{t \in K - \{i\}} (a_{r,it} + a_{ir,it} x_i) x_r x_t + \sum_{j \in K - \{i,r\}} \sum_{t \in K - \{i\}} (a_{j,it} + a_{ij,it} x_i) x_j x_t = \\ (a_{r,ir} + a_{ir,ir} x_i) x_r^2 &+ \sum_{j \in K - \{i,r\}} (a_{r,ij} + a_{j,ir} + 2 a_{ir,ij} x_i) x_j x_r + \sum_{j,t \in K - \{i,r\}} (a_{j,it} + a_{ij,it} x_i) x_j x_t, \\ \sum_{\substack{t,t' \in K - \{i\} \\ t < t'}} (a_{i,tt'} + \sum_{j \in K - \{i\}} a_{ij,tt'} x_j) x_t x_{t'} &= \sum_{t \in K - \{i,r\}} (a_{i,tr} + a_{ir,tr} x_r + \sum_{j \in K - \{i,r\}} a_{ij,tr} x_j) x_t x_r + \\ \sum_{\substack{t,t' \in K - \{i,r\} \\ t < t'}} (a_{i,tt'} + a_{ir,tt'} x_r + \sum_{j \in K - \{i,r\}} a_{ij,tt'} x_j) x_t x_{t'}. \quad \blacksquare \end{aligned}$$

Una posible elección del punto de inicio para el proceso iterativo del método de Newton-Raphson en el estadio $n+1$, sería considerar el punto *hibrido* elegido en el estadio anterior. Otra posible elección, sería considerar la elección *certainty equivalence*, aunque su cálculo precisa de la resolución de un sistema de ecuaciones equivalente, incluso para el estadio inicial. En cuanto a la elección del punto de inicio en el primer estadio se podría simplificar el número de parámetros del modelo y tomar sólo la parte de primer orden y obtener la solución a partir de la ecuación de la Proposición 6.1.

Apéndice 6.A.3 Demostración Lema 6.2 y Proposición 6.2.

Demostración Lema 6.2.

Utilizando los resultados del Lema 2.1 de inversión de matrices particionadas, se tiene que si

$$V_n[\mathbf{b}] = \mathbf{s}^2 (M + Z'_n Z_n)^{-1} = \begin{bmatrix} V_n[\mathbf{b}_0] & COV_n[\mathbf{b}_0, \mathbf{b}_I]' \\ COV_n[\mathbf{b}_0, \mathbf{b}_I] & V_n[\mathbf{b}_I] \end{bmatrix} \text{ y}$$

$$V_n^{-1}[\mathbf{b}] = \frac{(M + Z'_n Z_n)}{\mathbf{s}^2} = \frac{1}{\mathbf{s}^2} \begin{bmatrix} m_{00} + n & M'_{0I} + \mathbf{1}'_n X_n \\ M_{0I} + X'_n \mathbf{1}_n & M_{II} + X'_n X_n \end{bmatrix}$$

entonces

$$V_n^{-1}[\mathbf{b}_I] = \frac{1}{\mathbf{s}^2} \left(M_{II} + X'_n X_n - \frac{(M_{0I} + X'_n \mathbf{1}_n)(M'_{0I} + \mathbf{1}'_n X_n)}{m_{00} + n} \right),$$

$$V_n^{-1}[\mathbf{b}_I] COV_n[\mathbf{b}_0, \mathbf{b}_I] = -\frac{M_{0I} + X'_n \mathbf{1}_n}{m_{00} + n}.$$

Por otro lado,

$$E_n[\mathbf{b}] = (M + Z'_n Z_n)^{-1} (M\mathbf{m} + Z'_n Y_n) \Rightarrow E_n[\mathbf{b}_I] = \frac{1}{\mathbf{s}^2} (COV_n[\mathbf{b}_0, \mathbf{b}_I]; V_n[\mathbf{b}_I]) (M\mathbf{m} + Z'_n Y_n) =$$

$$\frac{1}{\mathbf{s}^2} (COV_n[\mathbf{b}_0, \mathbf{b}_I] (m_{00}\mathbf{m}_0 + M'_{0I}\mathbf{m}_I + \mathbf{1}'_n Y_n) + V_n[\mathbf{b}_I] (M_{0I}\mathbf{m}_0 + M_{II}\mathbf{m}_I + X'_n Y_n)),$$

con lo que

$$V_n^{-1}[\mathbf{b}_I] E_n[\mathbf{b}_I] = -\frac{1}{\mathbf{s}^2} \left(\frac{M_{0I} + X'_n \mathbf{1}_n}{m_{00} + n} (m_{00}\mathbf{m}_0 + M'_{0I}\mathbf{m}_I + \mathbf{1}'_n Y_n) - (M_{0I}\mathbf{m}_0 + M_{II}\mathbf{m}_I + X'_n Y_n) \right).$$

Sustituyendo en la Proposición 6.1, se tiene la expresión deseada. ■

Demostración Proposición 6.2.

Se procede por inducción sobre n . Asumamos que j es tal que $(m_{00}m'_j - m_{j0}M'_{0I})\mathbf{m}_I \neq 0$. Del

Lema 6.2 se obtiene que cuando $n=0$,

$$x_1^{hb}(H) = \frac{1}{m_{00}} M_{0I} + L_0(H) (M_{II} - \frac{1}{m_{00}} M_{0I} M'_{0I}) \mathbf{m}_I =$$

$$\frac{1}{m_{00}} (M_{0I} - m_{j0} e_j) + \frac{m_{j0}}{m_{00}} e_j + L_0(H) (m_{00}m'_j - m_{j0}M'_{0I}) \mathbf{m}_I e_j = \frac{M_{0I}}{m_{00}} + (x_{1,j}(H) - \frac{m_{j0}}{m_{00}}) e_j.$$

Supongamos cierto hasta el estadio n , entonces $x_i^{hb}(H) = (1/m_{00})M_{0I} + (x_{i,j}(H) - m_{j0}/m_{00})e_j$ para todo $i \leq n$. Si definimos x_j^n como la j -ésima columna de X_n , $x_j^n = (x_{1,j}(H), \dots, x_{n,j}(H))'$, esta hipótesis de inducción implica que $X_n' = (1/m_{00})M_{0I}1_n' + e_j(x_j^n - 1_n m_{j0}/m_{00})'$ y, por tanto,

$$X_n'1_n = (n/m_{00})M_{0I} + (1_n'x_j^n - n m_{j0}/m_{00})e_j$$

y

$$X_n'Y_n = 1_n'Y_n M_{0I}/m_{00} + (Y_n'x_j^n - m_{j0}/m_{00})e_j.$$

Del Lema 6.2 se tiene que

$$x_{n+1}^{hb} = \frac{1}{m_{00} + n}(M_{0I} + X_n'1_n) +$$

$$L_n(H) \left\{ M_{0I}\mathbf{m}_0 + M_{II}\mathbf{m}_I + X_n'Y_n - \frac{m_{00}\mathbf{m}_0 + M_{0I}'\mathbf{m}_I + 1_n'Y_n}{m_{00} + n}(M_{0I} + X_n'1_n) \right\},$$

y de aquí

$$x_{n+1,j}(H) = \frac{m_{j0} + 1_n'x_j^n}{m_{00} + n} + L_n(H) \left\{ m_{j0}\mathbf{m}_0 + m_j'\mathbf{m}_I + Y_n'x_j^n - \frac{m_{00}\mathbf{m}_0 + M_{0I}'\mathbf{m}_I + 1_n'Y_n}{m_{00} + n}(m_{j0} + 1_n'x_j^n) \right\}.$$

Así que

$$\begin{aligned} \frac{M_{0I}}{m_{00}} + (x_{n+1,j}(H) - \frac{m_{j0}}{m_{00}})e_j &= x_{n+1}^{hb}(H) + \frac{M_{0I}}{m_{00}} - \frac{m_{j0}}{m_{00}}e_j + \\ &\quad \frac{1}{m_{00} + n}(-M_{0I} - n\frac{M_{0I}}{m_{00}} + (m_{j0} + n\frac{m_{j0}}{m_{00}})e_j) + \\ &\quad L_n(H) \{ (m_{j0}\mathbf{m}_0 + m_j'\mathbf{m}_I)e_j - M_{0I}\mathbf{m}_0 - M_{II}\mathbf{m}_I - 1_n'Y_n\frac{M_{0I}}{m_{00}} + \\ &\quad \frac{m_{j0}}{m_{00}}1_n'Y_n e_j - \frac{m_{00}\mathbf{m}_0 + M_{0I}'\mathbf{m}_I + 1_n'Y_n}{m_{00} + n}(m_{j0}e_j - M_{0I} - n\frac{M_{0I}}{m_{00}} + n\frac{m_{j0}}{m_{00}}e_j) \}. \end{aligned}$$

La primera parte de la demostración se obtiene chequeando que

$$\frac{1}{m_{00} + n}(-M_{0I} - n\frac{M_{0I}}{m_{00}} + (m_{j0} + n\frac{m_{j0}}{m_{00}})e_j) = -(\frac{M_{0I}}{m_{00}} - \frac{m_{j0}}{m_{00}}e_j)$$

y

$$\begin{aligned}
& (m_{j_0} \mathbf{m}_0 + m'_j \mathbf{m}_l) e_j - M_{0l} \mathbf{m}_0 - M_{ll} \mathbf{m}_l - 1'_n Y_n \frac{M_{0l}}{m_{00}} + \\
& \frac{m_{j_0}}{m_{00}} 1'_n Y_n e_j - \frac{m_{00} \mathbf{m}_0 + M'_{0l} \mathbf{m}_l + 1'_n Y_n}{m_{00} + n} (m_{j_0} e_j - M_{0l} - n \frac{M_{0l}}{m_{00}} + n \frac{m_{j_0}}{m_{00}} e_j) = \\
& m'_j \mathbf{m}_l e_j - M_{ll} \mathbf{m}_l + M'_{0l} \mathbf{m}_l \frac{M_{0l}}{m_{00}} - M'_{0l} \mathbf{m}_l \frac{m_{j_0}}{m_{00}} e_j = \\
& (m'_j - \frac{m_{j_0}}{m_{00}} M'_{0l}) \mathbf{m}_l e_j - (M_{ll} - \frac{M_{0l}}{m_{00}} M'_{0l}) \mathbf{m}_l = 0.
\end{aligned}$$

Cuando no existe $j \in \{1, \dots, k\}$ tal que $(m_{00} m'_j - m_{j_0} M'_{0l}) \mathbf{m}_l \neq 0$, se verifica que $(m_{00} M_{ll} - M_{0l} M'_{0l}) \mathbf{m}_l = 0$. Procediendo igualmente por inducción, el resultado se verifica trivialmente para $n=0$. Supuesto cierto para $i \leq n$, entonces $X'_n 1_n = (n/m_{00}) M_{0l}$ y $X'_n Y_n = (1'_n Y_n / m_{00}) M_{0l}$ y, sustituyendo en el Lema 6.2, se obtiene que $x_{n+1}^{hb}(H) = (1/m_{00}) M_{0l}$.

■

Apéndice 6.A.4 Cálculo de $x_n^{hb}(H)$ para un modelo lineal bajo la restricción de que $C = B(R)$.

En este apéndice se introduce el método de Newton-Raphson usado para encontrar el punto que minimiza el $hb_{n+1}(H)$ bajo las restricciones de que $x_n^{hb}(H) \in C$ y para un modelo lineal con interacciones donde C es una bola de radio R centrada en el origen. Se podrían utilizar otros métodos de minimización o de resolución de sistemas no lineales, pero el método de Newton-Raphson funciona bien para este problema ya que el cálculo de la matriz jacobiana es trivial. A continuación se describe dicho método. Consideremos el problema en el que se dan N relaciones funcionales igualadas a cero

$$F_i(t_1, t_2, \dots, t_N) = 0, \quad \text{con } i=1, 2, \dots, N \text{ y } t_i \in R.$$

Sea $t = (t_1, t_2, \dots, t_N)$ y $F = (F_1, F_2, \dots, F_N)$. En las proximidades de t , cada una de las F_i puede ser expandida en serie de Taylor como

$$F_i(t + \mathbf{d} t) = F_i(t) + \sum_{j=1}^N \frac{\partial F_i}{\partial t_j} \partial t_j + O(\partial t^2).$$

Denotando la matriz jacobiana con elementos

$$J_{ij} \equiv \frac{\partial F_i}{\partial t_j}$$

por J , se tiene en notación matricial que

$$F(t + \partial t) = F(t) + J \partial t + O(\partial t^2).$$

Despreciando los términos de orden mayor o igual a dos, e igualando $F(t + \partial t) = 0$, se obtiene un conjunto de ecuaciones lineales para la corrección ∂t igual a

$$J \partial t = -F,$$

y que puede ser resuelta usando la descomposición LU . La corrección será añadida al vector solución

$$t_{nuevo} = t_{viejo} + \partial t$$

y el proceso será iterado hasta la convergencia.

Para nuestro caso particular de un modelo lineal de primer orden, se debe resolver el siguiente problema de minimización:

$$\begin{cases} \min h b_{n+1}(H) \\ \text{s.a. } \sum x_i^2 \leq R^2 \end{cases}.$$

Si el punto que minimiza $h b_{n+1}(H)$ cae fuera de la región $C = B(R)$ esto se traduce en

$$\left. \begin{array}{l} \min z'Az - 2T z'E_n[\mathbf{b}] \\ \text{s.a. } \sum x_i^2 = R^2 \end{array} \right\} \Leftrightarrow \min z'Az - 2T z'E_n[\mathbf{b}] + \mathbf{I}(\sum x_i^2 - R^2),$$

donde \mathbf{I} es el multiplicador de Lagrange asociado a tal restricción, $A = E_n[\mathbf{b}]E_n[\mathbf{b}]' + H V_n[\mathbf{b}]$ es una matriz $(k+1) \times (k+1)$ y $z' = (1 | x') = (1, x_1, \dots, x_k)$. Derivando con respecto a x y \mathbf{I} se obtiene el siguiente sistema no lineal (las funciones F_i del método de Newton-Raphson):

$$\begin{cases} (A_{II} + \mathbf{I} I_k)x + (A_{0I} - T E_n[\mathbf{b}_I]) = 0 \\ x'x - R^2 = 0 \end{cases},$$

con la matriz A particionada como sigue

$$A = \begin{pmatrix} a_{00} & A'_{0I} \\ A_{0I} & A_{II} \end{pmatrix},$$

donde A_{II} es una matriz $k \times k$. La matriz jacobiana resultante para dicho sistema queda como

$$J = \begin{pmatrix} A_{II} + \mathbf{I} & I_k & x \\ 2x' & & 0 \end{pmatrix}.$$

Un punto inicial, $x^0 = (x_1^0, \dots, x_k^0)$, podría ser

$$x_i^0 = \sqrt{\frac{R^2}{\sum x_i^2}} x_i \quad \text{para } i = 1, \dots, k,$$

que coincide con el punto de la frontera de la región C que se obtiene de unir con una línea recta el punto del diseño *híbrido* sin restricciones, $x = (x_1, \dots, x_k)$, y el centro de la región C , que en este caso es el origen.

APENDICE I.

RESULTADOS DISTRIBUCIONALES: MODELO DE REGRESIÓN SIMPLE NORMAL CON PRIORI SOBRE b_0 y b_1 .

El modelo de regresión simple normal asume que $E[y|x, \mathbf{b}] = \mathbf{b}_0 + \mathbf{b}_1 x$ y $Var(\mathbf{e} | x, \mathbf{b}) = \mathbf{s}^2$ o, en forma matricial, que

$$Y_n | X_n, \mathbf{b} \sim N_n(Z_n \mathbf{b}, \mathbf{s}^2 I_n),$$

con Z_n la matriz de $n \times 2$ cuya j -ésima fila es el vector $z'_j = (1|x_j) = (1, x_j)$, $Y'_n = (y_1, y_2, \dots, y_n)$, $\mathbf{b}' = (\mathbf{b}_0, \mathbf{b}_1)$ e I_n es la matriz identidad $n \times n$. En un contexto bayesiano, se supone una distribución a priori sobre dichos parámetros que identificaremos por $\mathbf{p}_0(\mathbf{b} | \mathbf{s})$. La elección de dicha distribución dependerá de la información inicial de la que se parta sobre el problema.

La clave estadística del modelo lineal normal se basa en:

1. definir las variables x e y (posiblemente usando transformaciones) para que la esperanza condicional de y sea razonablemente lineal como una función de x con errores aproximadamente normales y,
2. fijar una distribución a priori sobre los parámetros del modelo que refleje exactamente el conocimiento inicial; una distribución a priori que sea suficientemente informativa para los parámetros del modelo considerado siempre que no domine a los datos de forma inapropiada.

El modelo lineal ordinario adopta varias suposiciones, linealidad de la esperanza $E[y|x, \mathbf{b}]$ como función de x , normalidad de los términos de error y observaciones independientes con igual varianza, ninguna de las cuales se cumple en la práctica. La pregunta clara es si hay demasiada diferencia entre la teoría y la práctica. Existen gran cantidad de métodos para chequear cómo ajustar el modelo a los datos y siempre se pueden probar distintos modelos para ver la sensibilidad de las inferencias.

En este primer apéndice se consideran cuatro distribuciones a priori, obtenidas suponiendo \mathbf{b}_0 y \mathbf{b}_1 independientes, y combinando la distribución uniforme y normal sobre ambos parámetros y bajo la suposición de que \mathbf{s} es considerada conocida. Para cada una de las cuatro distribuciones a priori se calculan las distribuciones a posteriori, las condicionales a posteriori de un parámetro sobre el otro y las marginales a posteriori para cada parámetro, siempre utilizando la notación dada en la Sección 1.7 del primer capítulo. Además, se calcula la moda a posteriori, que en el caso de priori conjugada coincide con la media a posteriori, y la varianza de predicción resultante de sustituir estos estimadores en el modelo. Estos cálculos proporcionan algunos de los ingredientes necesarios para implementar de forma eficiente las familias de diseños introducidas en los capítulos de esta tesis.

La primera priori que se tratará es aquella en la que \mathbf{b}_0 y \mathbf{b}_1 son normales, no necesariamente independientes, correspondiente al caso conjugado. El hecho de que los momentos de la distribución a posteriori de los parámetros se puedan dar de forma cerrada hace a esta priori una de las más empleadas en las distintas ejemplificaciones de las familias de diseños implementadas en esta tesis. Como última sección de este apéndice se incluyen resultados sobre estimadores mínimos cuadrados que son de gran utilidad en el desarrollo de éste.

1. PRIORI CONJUGADA.

Consideremos que $\mathbf{p}_0(\mathbf{b} | \mathbf{s})$ es $N_2(\mathbf{m}, \mathbf{s}^2 M^{-1})$, con $\mathbf{m}' = (\mathbf{m}_{b_0}, \mathbf{m}_{b_1})$ y $M = \{m_{ij}\}$ una matriz de 2×2 . La densidad a priori es

$$\mathbf{p}_0(\mathbf{b}) = \frac{1}{(2\mathbf{p})^{2/2} |\mathbf{s}^2 M^{-1}|^{1/2}} \exp\left\{-\frac{(\mathbf{b} - \mathbf{m})' (\mathbf{s}^2 M^{-1})^{-1} (\mathbf{b} - \mathbf{m})}{2}\right\} \propto \exp\left\{-\frac{(\mathbf{b} - \mathbf{m})' M (\mathbf{b} - \mathbf{m})}{2\mathbf{s}^2}\right\},$$

y la verosimilitud es

$$f(Y_n | X_n, \mathbf{b}) = \frac{1}{(2\mathbf{p})^{n/2} |\mathbf{s}^2 I_n|^{1/2}} \exp\left\{-\frac{(Y_n - Z_n \mathbf{b})' (\mathbf{s}^2 I_n)^{-1} (Y_n - Z_n \mathbf{b})}{2}\right\} \\ \propto \exp\left\{-\frac{(Y_n - Z_n \mathbf{b})' (Y_n - Z_n \mathbf{b})}{2\mathbf{s}^2}\right\},$$

con lo que la densidad conjunta para Y_n y \mathbf{b} viene dada como

$$h(Y_n, \mathbf{b} | X_n, \mathbf{p}_0) = f(Y_n | X_n, \mathbf{b}) \mathbf{p}_0(\mathbf{b}) \propto \exp\left\{-\frac{(Y_n' Y_n + \mathbf{m}' M \mathbf{m}) + \mathbf{b}' (M + Z_n' Z_n) \mathbf{b} - 2\mathbf{b}' (M \mathbf{m} + Z_n' Y_n)}{2\mathbf{s}^2}\right\}.$$

a) Conjunta a posteriori.

$$\mathbf{p}_n(\mathbf{b}) \propto \exp\left\{-\frac{(\mathbf{b} - (M + Z_n' Z_n)^{-1} (M \mathbf{m} + Z_n' Y_n))' (M + Z_n' Z_n) (\mathbf{b} - (M + Z_n' Z_n)^{-1} (M \mathbf{m} + Z_n' Y_n))}{2\mathbf{s}^2}\right\}.$$

De aquí,

$$\mathbf{p}_n(\mathbf{b}) \sim N_2\left\{(M + Z_n' Z_n)^{-1} (M \mathbf{m} + Z_n' Y_n), \mathbf{s}^2 (M + Z_n' Z_n)^{-1}\right\}$$

y denotando $\mathbf{z}' = (1, x)$ se tiene que

$$\mathbf{p}_n(\mathbf{b}_0 + \mathbf{b}_1 x) \sim N\left\{\mathbf{z}' (M + Z_n' Z_n)^{-1} (M \mathbf{m} + Z_n' Y_n), \mathbf{s}^2 \mathbf{z}' (M + Z_n' Z_n)^{-1} \mathbf{z}\right\}.$$

b) Condicionales a posteriori.

$$\mathbf{p}_n(\mathbf{b}_0 | \mathbf{b}_1) \sim N\left(\frac{m_{11}\mathbf{m}_{b_0} + \sum y_i - \mathbf{b}_1 \sum x_i + m_{12}(\mathbf{m}_{b_1} - \mathbf{b}_1)}{m_{11} + n}, \frac{\mathbf{S}^2}{m_{11} + n}\right)$$

y

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{b}_0) \sim N\left(\frac{m_{22}\mathbf{m}_{b_1} + \sum x_i y_i - \mathbf{b}_0 \sum x_i + m_{12}(\mathbf{m}_{b_0} - \mathbf{b}_0)}{m_{22} + \sum x_i^2}, \frac{\mathbf{S}^2}{m_{22} + \sum x_i^2}\right).$$

Éstas se obtienen desarrollando la parte del exponente de $\mathbf{p}_n(\mathbf{b})$ que depende de β , es decir,

$$\begin{aligned} & -\frac{\mathbf{b}_0^2(m_{11} + n) + \mathbf{b}_1^2(m_{22} + \sum x_i^2) + 2\mathbf{b}_0\mathbf{b}_1(m_{12} + \sum x_i) - 2\mathbf{b}_0(m_{11}\mathbf{m}_{b_0} + m_{12}\mathbf{m}_{b_1} + \sum y_i)}{2\mathbf{S}^2} \\ & + \frac{2\mathbf{b}_1(m_{12}\mathbf{m}_{b_0} + m_{22}\mathbf{m}_{b_1} + \sum x_i y_i)}{2\mathbf{S}^2}. \end{aligned}$$

c) Marginales a posteriori.

Se construyen de forma inmediata sin más que tener en cuenta, que si $\mathbf{p}_n(\mathbf{b}) \sim N_2(\mathbf{m}, \mathbf{S})$ entonces $\mathbf{p}_n(C'\mathbf{b}) \sim N(C'\mathbf{m}, C'\mathbf{S}C)$; en particular, la marginal de \mathbf{b}_0 y \mathbf{b}_1 se obtiene para $C' = (1,0)$ y $C' = (0,1)$, respectivamente.

d) Modas a posteriori.

Para esta priori coinciden con las medias a posteriori y son

$$\begin{aligned} E_n[\mathbf{b}_0] &= \frac{(\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i) + m_{22}(\sum y_i - \mathbf{m}_1 \sum x_i) + m_{11}(m_{22} + \sum x_i^2) \mathbf{m}_{b_0}}{w_{1n} + w_{2n} + w_{3n} + w_{4n}} \\ &+ \frac{m_{12}(\mathbf{m}_{b_1} \sum x_i^2 - \sum x_i y_i - \mathbf{m}_{b_0}(m_{12} + \sum x_i))}{w_{1n} + w_{2n} + w_{3n} + w_{4n}} = \\ & \frac{w_{1n} \hat{\mathbf{b}}_{0n}^{ls} + w_{2n}(\bar{y}_n - \mathbf{m}_{b_1} \bar{x}_n) + w_{3n} \mathbf{m}_{b_0} + m_{12}(\mathbf{m}_{b_1} \sum x_i^2 - \sum x_i y_i - \mathbf{m}_{b_0}(m_{12} + \sum x_i))}{w_{1n} + w_{2n} + w_{3n} + w_{4n}}, \end{aligned}$$

$$E_n[\mathbf{b}_1] = \frac{(n \sum x_i y_i - \sum x_i \sum y_i) + n m_{22} \mathbf{m}_{b_1} + m_{11} (m_{22} \mathbf{m}_{b_1} - \mathbf{m}_{b_0} \sum x_i + \sum x_i y_i)}{w_{1n} + w_{2n} + w_{3n} + w_{4n}}$$

$$+ \frac{m_{12} (n \mathbf{m}_{b_0} - \sum y_i - \mathbf{m}_{b_1} (m_{12} + \sum x_i))}{w_{1n} + w_{2n} + w_{3n} + w_{4n}} =$$

$$\frac{w_{1n} \hat{\mathbf{b}}_{1n}^{ls} + w_{2n} \mathbf{m}_{b_1} + m_{11} (m_{22} \mathbf{m}_{b_1} - \mathbf{m}_{b_0} \sum x_i + \sum x_i y_i) + m_{12} (n \mathbf{m}_{b_0} - \sum y_i - \mathbf{m}_{b_1} (m_{12} + \sum x_i))}{w_{1n} + w_{2n} + w_{3n} + w_{4n}},$$

donde $w_{1n} = n \sum (x_i - \bar{x}_n)^2 = |Z'_n Z_n|$, $w_{2n} = n m_{22}$, $w_{3n} = m_{11} (m_{22} + \sum x_i^2)$ y $w_{4n} = -m_{12} (m_{12} + 2 \sum x_i)$.

Cuando \mathbf{b}_0 y \mathbf{b}_1 son independientes a priori ($m_{12} = 0$), las medias a posteriori se pueden expresar como

$$E_n[\mathbf{b}_0] = \frac{w_{1n} \hat{\mathbf{b}}_{0n}^{ls} + w_{2n} (\bar{y}_n - \mathbf{m}_{b_1} \bar{x}_n) + w_{3n} \mathbf{m}_{b_0}}{w_{1n} + w_{2n} + w_{3n}},$$

$$E_n[\mathbf{b}_1] = \frac{w_{1n} \hat{\mathbf{b}}_{1n}^{ls} + w_{2n} \mathbf{m}_{b_1} + w_{3n} (\mathbf{m}_{b_1} + \frac{\sum x_i y_i - \mathbf{m}_{b_0} \sum x_i}{\sum x_i^2}) - m_{11} (\mathbf{m}_{b_1} \sum x_i^2 + m_{22} \frac{\sum x_i y_i - \mathbf{m}_{b_0} \sum x_i}{\sum x_i^2})}{w_{1n} + w_{2n} + w_{3n}}.$$

e) Varianzas y Covarianzas de las modas a posteriori.

En notación matricial se tiene que

$$Var _ Cov_n(E_n[\mathbf{b}_0], E_n[\mathbf{b}_1]) = \mathbf{s}^2 (M + Z'_n Z_n)^{-1} Z'_n Z_n (M + Z'_n Z_n)^{-1}.$$

Para implementación computacional se puede utilizar que

$$V_n(E_n[\mathbf{b}_0]) = \mathbf{s}^2 \frac{w_{1n} \sum x_i^2 + m_{22} (w_{2n} + 2w_{1n}) + m_{12} (m_{11} \sum x_i^2 - 2m_{22} \sum x_i)}{(w_{1n} + w_{2n} + w_{3n} + w_{4n})^2},$$

$$V_n(E_n[\mathbf{b}_1]) = \mathbf{s}^2 \frac{n w_{1n} + m_{11} (m_{11} \sum x_i^2 + 2w_{1n}) + m_{12} (n m_{12} - 2m_{11} \sum x_i)}{(w_{1n} + w_{2n} + w_{3n} + w_{4n})^2},$$

$$\text{Cov}_n(E_n[\mathbf{b}_0], E_n[\mathbf{b}_1]) = \mathbf{s}^2 \frac{\sum x_i (m_{11}m_{22} - w_{1n}) - m_{12}(w_{2n} + 2w_{1n} + m_{11} \sum x_i^2 - m_{12} \sum x_i)}{(w_{1n} + w_{2n} + w_{3n} + w_{4n})^2}.$$

f) Varianza de estimación de predicción, $E_n[y(x_0)] = E_n[\mathbf{b}_0 + \mathbf{b}_1 x_0]$.

$$V_n(E_n[y(x)]) = x_0^2 V_n(E_n[\mathbf{b}_1]) + 2x_0 \text{Cov}_n(E_n[\mathbf{b}_0], E_n[\mathbf{b}_1]) + V_n(E_n[\mathbf{b}_0]).$$

Por otro lado, la varianza de $f(x; \mathbf{b})$ es

$$V_n(f(x; \mathbf{b})) = \frac{x_0^2 (n + m_{11}) - 2x_0 (m_{12} + \sum x_i) + (m_{22} + \sum x_i^2)}{w_{1n} + w_{2n} + w_{3n} + w_{4n}} \mathbf{s}^2.$$

2. PRIORI NORMAL SOBRE \mathbf{b}_0 Y UNIFORME SOBRE \mathbf{b}_1 .

Consideremos $\mathbf{p}_0(\mathbf{b})$ tal que $\mathbf{b}_0 \sim N(\mathbf{m}_{b_0}, \mathbf{s}_{b_0}^2)$ y $\mathbf{b}_1 \sim U[lb_1, upb_1]$ independientes. La densidad a priori es

$$\mathbf{p}_0(\mathbf{b}_0, \mathbf{b}_1) \propto I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{-\frac{1}{2}\left(\frac{\mathbf{b}_0 - \mathbf{m}_{b_0}}{\mathbf{s}_{b_0}}\right)^2\right\}.$$

La distribución conjunta de Y_n y \mathbf{b} queda como

$$h(Y_n, \mathbf{b} | X_n, \mathbf{p}_0) \propto I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{-\frac{1}{2}\left(\frac{(Y_n - Z_n \mathbf{b})'(Y_n - Z_n \mathbf{b})}{\mathbf{s}^2} + \frac{(\mathbf{b}_0 - \mathbf{m}_{b_0})^2}{\mathbf{s}_{b_0}^2}\right)\right\}.$$

a) Conjunta a posteriori.

$$\mathbf{p}_n(\mathbf{b}) \propto I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{-\frac{1}{2\mathbf{s}^2 \mathbf{s}_{b_0}^2} (\mathbf{s}_{b_0}^2 (\mathbf{b}' Z_n' Z_n \mathbf{b} - 2Y_n' Z_n \mathbf{b}) + \mathbf{s}^2 (\mathbf{b}_0^2 - 2\mathbf{b}_0 \mathbf{m}_{b_0}))\right\}.$$

b) Condicionales a posteriori.

$$\mathbf{p}_n(\mathbf{b}_0 | \mathbf{b}_1) \sim \begin{cases} N\left(\frac{\mathbf{s}_{b_0}^2 (\sum y_i - \mathbf{b}_1 \sum x_i)}{n\mathbf{s}_{b_0}^2 + \mathbf{s}^2}, \frac{\mathbf{s}^2 \mathbf{s}_{b_0}^2}{n\mathbf{s}_{b_0}^2 + \mathbf{s}^2}\right) & \text{si } \mathbf{b}_1 \in [lb_1, upb_1] \\ 0 & \text{en el resto.} \end{cases}$$

Cuando $\sum x_i^2 \neq 0$,

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{b}_0) \sim N^{[lb_1, upb_1]}\left(\mathbf{s}_{b_0}^2 \frac{\sum x_i y_i - \mathbf{b}_0 \sum x_i}{\sum x_i^2}, \frac{\mathbf{s}^2}{\sum x_i^2}\right),$$

y si $\sum x_i^2 = 0$,

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{b}_0) \sim U[lb_1, upb_1].$$

Éstas se obtienen desarrollando la parte del exponente de $\mathbf{p}_n(\mathbf{b})$ que depende de \mathbf{b} , es decir,

$$\frac{\mathbf{b}_0^2 (n\mathbf{s}_{b_0}^2 + \mathbf{s}^2) + 2\mathbf{b}_0 (\mathbf{s}_{b_0}^2 (\mathbf{b}_1 \sum x_i - \sum y_i) - \mathbf{s}^2 \mathbf{m}_{b_0}) + \mathbf{s}_{b_0}^2 (\mathbf{b}_1^2 \sum x_i^2 - 2\mathbf{b}_1 \sum x_i y_i)}{2\mathbf{s}^2 \mathbf{s}_{b_0}^2}.$$

c) Marginales a posteriori.

Cuando $\sum x_i^2 \neq 0$ se tiene que

$$\mathbf{p}_n(\mathbf{b}_1) \sim N^{[lb_1, upb_1]} \left(\frac{w_{1n} \hat{\mathbf{b}}_{1n}^{ls} + w_{2n} (\sum x_i y_i - \mathbf{m}_{b_0} \sum x_i) / \sum x_i^2}{w_{1n} + w_{2n}}, \frac{(n\mathbf{s}_{b_0}^2 + \mathbf{s}^2) \mathbf{s}^2}{w_{1n} + w_{2n}} \right)$$

y

$$\mathbf{p}_n(\mathbf{b}_0) \propto \left[\mathbf{F} \left(\frac{upb_1 - E_{b_0}^*}{\sqrt{V_{b_0}^*}} \right) - \mathbf{F} \left(\frac{lb_1 - E_{b_0}^*}{\sqrt{V_{b_0}^*}} \right) \right] \exp \left\{ \frac{1}{2} \left[\frac{(E_{b_0}^*)^2}{V_{b_0}^*} - \frac{\mathbf{b}_0^2 (n\mathbf{s}_{b_0}^2 + \mathbf{s}^2) - 2\mathbf{b}_0 (\mathbf{s}_{b_0}^2 \sum y_i + \mathbf{s}^2 \mathbf{m}_{b_0})}{\mathbf{s}^2 \mathbf{s}_{b_0}^2} \right] \right\},$$

con $w_{1n} = n\mathbf{s}_{b_0}^2 \sum (x_i - \bar{x}_n)^2 = \mathbf{s}_{b_0}^2 |Z_n' Z_n|$, $w_{2n} = \mathbf{s}^2 \sum x_i^2$, $E_{b_0}^* = \mathbf{s}_{b_0}^2 (\sum x_i y_i - \mathbf{b}_0 \sum x_i) / \sum x_i^2$ y $V_{b_0}^* = \mathbf{s}^2 / \sum x_i^2$.

Si $\sum x_i^2 = 0$ entonces

$$\mathbf{p}_n(\mathbf{b}_1) \sim U(lb_1, upb_1)$$

y

$$\mathbf{p}_n(\mathbf{b}_0) \sim N \left(\frac{\mathbf{s}_{b_0}^2 \sum y_i + \mathbf{s}^2 \mathbf{m}_{b_0}}{n\mathbf{s}_{b_0}^2 + \mathbf{s}^2}, \frac{\mathbf{s}^2 \mathbf{s}_{b_0}^2}{n\mathbf{s}_{b_0}^2 + \mathbf{s}^2} \right).$$

d) Modas a posteriori.

Si $\sum x_i^2 \neq 0$, \mathbf{b}_{1n}^{pmo} es el punto en $[lb_1, upb_1]$ que está más cerca de

$$\hat{\mathbf{b}}_{1n} = \frac{\mathbf{s}_{b_0}^2 (n \sum x_i y_i - \sum x_i \sum y_i) + \mathbf{s}^2 (\sum x_i y_i - \mathbf{m}_{b_0} \sum x_i)}{n\mathbf{s}_{b_0}^2 \sum (x_i - \bar{x}_n)^2 + \mathbf{s}^2 \sum x_i^2}.$$

Si $\hat{\mathbf{b}}_{1n} \in [lb_1, upb_1]$,

$$\mathbf{b}_{0n}^{pmo} = \frac{\mathbf{s}_{b_0}^2 (\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i) + \mathbf{s}^2 \mathbf{m}_{b_0} \sum x_i}{n \mathbf{s}_{b_0}^2 \sum (x_i - \bar{x}_n)^2 + \mathbf{s}^2 \sum x_i^2},$$

mientras que si $\hat{\mathbf{b}}_{1n} \notin [lb_1, upb_1]$, \mathbf{b}_{1n}^{pmo} es el extremo del intervalo más próximo a $\hat{\mathbf{b}}_{1n}$, y si éste es lb_1 ,

$$\mathbf{b}_{0n}^{pmo} = \frac{\mathbf{s}_{b_0}^2 (\sum y_i - lb_1 \sum x_i) + \mathbf{s}^2 \mathbf{m}_{b_0}}{n \mathbf{s}_{b_0}^2 + \mathbf{s}^2}$$

(de forma análoga cuando $\mathbf{b}_{1n}^{pmo} = upb_1$).

Si $\sum x_i^2 = 0$, \mathbf{b}_{1n}^{pmo} es cualquier punto del intervalo $[lb_1, upb_1]$ y

$$\mathbf{b}_{0n}^{pmo} = \frac{\mathbf{s}_{b_0}^2 \sum y_i + \mathbf{s}^2 \mathbf{m}_{b_0}}{n \mathbf{s}_{b_0}^2 + \mathbf{s}^2}.$$

Estos estimadores se obtienen maximizando el exponente de la densidad $\mathbf{p}_n(\mathbf{b})$ con la restricción de que $\mathbf{b}_1 \in (lb_1, upb_1)$, lo que lleva a resolver el sistema

$$\begin{cases} \mathbf{b}_0 (n \mathbf{s}_{b_0}^2 + \mathbf{s}^2) + \mathbf{s}_{b_0}^2 (\mathbf{b}_1 \sum x_i - \sum y_i) - \mathbf{s}^2 \mathbf{m}_{b_0} = 0 \\ \mathbf{b}_0 \sum x_i + \mathbf{b}_1 \sum x_i^2 - \sum x_i y_i = 0. \end{cases}$$

e) Varianzas y Covarianzas de las modas a posteriori.

Cuando $\sum x_i^2 \neq 0$,

$$\begin{aligned} V_n(\mathbf{b}_{1n}^{pmo} | \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) &= \frac{1}{(w_{1n} + w_{2n})^2} V_n(w_{1n} \hat{\mathbf{b}}_{1n}^{ls} + \mathbf{s}^2 (\sum x_i y_i - \mathbf{m}_{b_0} \sum x_i)) = \\ &= \frac{1}{(w_{1n} + w_{2n})^2} \left(w_{1n}^2 \frac{\mathbf{s}^2}{\sum (x_i - \bar{x}_n)^2} + \mathbf{s}^4 \mathbf{s}^2 \sum x_i^2 + 2w_{1n} \mathbf{s}^2 \text{Cov}_n(\hat{\mathbf{b}}_{1n}^{ls}, \sum x_i y_i) \right) = \\ &= \frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \left(\frac{w_{1n}^2}{\sum (x_i - \bar{x}_n)^2} + \mathbf{s}^2 (w_{2n} + 2w_{1n}) \right). \end{aligned}$$

$$V_n(\mathbf{b}_{0n}^{pmo} | \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) = \frac{V_n(w_{1n} \hat{\mathbf{b}}_{0n}^{ls} + w_{2n} \mathbf{m}_{b_0})}{(w_{1n} + w_{2n})^2} = \frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \frac{w_{1n}^2 \sum x_i^2}{n \sum (x_i - \bar{x}_n)^2},$$

$$\text{Cov}_n(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo} | \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) = \frac{w_{1n}^2 \text{Cov}_n(\hat{\mathbf{b}}_{0n}^{ls}, \hat{\mathbf{b}}_{1n}^{ls}) + w_{1n} \mathbf{s}^2 \text{Cov}_n(\hat{\mathbf{b}}_{0n}^{ls}, \sum x_i y_i)}{(w_{1n} + w_{2n})^2} =$$

$$\frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \left(\frac{-w_{1n}^2 \sum x_i}{n \sum (x_i - \bar{x}_n)^2} \right).$$

Según esto, la matriz de varianzas covarianzas viene dada por

$$\text{Var} - \text{Cov}_n(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo} | \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) =$$

$$\frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \left(w_{1n}^2 (Z_n' Z_n)^{-1} + (2w_{1n} w_{2n} + w_{2n}^2) \begin{pmatrix} 0 & 0 \\ 0 & 1/\sum x_i^2 \end{pmatrix} \right),$$

con $w_{1n} = n \mathbf{s}_{b_0}^2 \sum (x_i - \bar{x}_n)^2 = \mathbf{s}_{b_0}^2 |Z_n' Z_n|$ y $w_{2n} = \mathbf{s}^2 \sum x_i^2$.

Si $\sum x_i^2 = 0$ ó $\mathbf{b}_{1n}^{pmo} \notin (lb_1, upb_1)$,

$$V_n(\mathbf{b}_{0n}^{pmo}) = \frac{\mathbf{s}^2 \mathbf{s}_{b_0}^4}{(\mathbf{s}_{b_0}^2 n + \mathbf{s}^2)^2}.$$

f) Varianza de estimación de predicción, $f_n^{pm}(x_0; \mathbf{b}) = \mathbf{b}_{0n}^{pmo} + \mathbf{b}_{1n}^{pmo} x_0$.

Cuando $\sum x_i^2 \neq 0$,

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) = \frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \left[w_{1n}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x}_n)^2}{\sum (x_i - \bar{x}_n)^2} \right) + \frac{(w_{2n}^2 + 2w_{1n} w_{2n}) x_0^2}{\sum x_i^2} \right],$$

con $w_{1n} = n \mathbf{s}_{b_0}^2 \sum (x_i - \bar{x}_n)^2 = \mathbf{s}_{b_0}^2 |Z_n' Z_n|$ y $w_{2n} = \mathbf{s}^2 \sum x_i^2$, y

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{b}_{1n}^{pmo} = lb_1) = \frac{\mathbf{s}^2 \mathbf{s}_{b_0}^4}{(\mathbf{s}_{b_0}^2 n + \mathbf{s}^2)^2} = V_n(\mathbf{b}_{0n}^{pmo} | \mathbf{b}_{1n}^{pmo} = lb_1).$$

De forma análoga se obtiene para $\mathbf{b}_{1n}^{pmo} = upb_1$.

Si $\sum x_i^2 = 0$,

$$V_n(f_n^{pm}(x_0; \mathbf{b})) = \frac{\mathbf{s}^2 \mathbf{s}_{b_0}^4}{(\mathbf{s}_{b_0}^2 n + \mathbf{s}^2)^2} = V_n(\mathbf{b}_{0n}^{pmo} | \sum x_i^2 = 0).$$

Notar que

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) =$$

$$\frac{w_{1n}^2 V_n(\hat{\mathbf{b}}_{0n}^{ls} + x_0 \hat{\mathbf{b}}_{1n}^{ls}) + (w_{2n}^2 + 2w_{1n} w_{2n}) V_n(\mathbf{m}_{b_0} + x_0 (\sum x_i y_i - \mathbf{m}_{b_0} \sum x_i) / \sum x_i^2)}{(w_{1n} + w_{2n})^2}.$$

3. PRIORI NORMAL SOBRE \mathbf{b}_1 Y UNIFORME SOBRE \mathbf{b}_0 .

Consideremos $\mathbf{p}_0(\mathbf{b})$ tal que $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1}, \mathbf{s}_{b_1}^2)$, $\mathbf{b}_0 \sim U[lb_0, upb_0]$ e independientes.

La densidad a priori es

$$\mathbf{p}_0(\mathbf{b}) \propto I_{[lb_0, upb_0]}(\mathbf{b}_0) \exp\left\{-\frac{1}{2}\left(\frac{\mathbf{b}_1 - \mathbf{m}_{b_1}}{\mathbf{s}_{b_1}}\right)^2\right\}.$$

La distribución conjunta de Y_n y \mathbf{b} queda como sigue

$$h(Y_n, \mathbf{b} | X_n, \mathbf{p}_0) \propto I_{[lb_0, upb_0]}(\mathbf{b}_0) \exp\left\{-\frac{1}{2}\left(\frac{(Y_n - Z_n \mathbf{b})'(Y_n - Z_n \mathbf{b})}{\mathbf{s}^2} + \frac{(\mathbf{b}_1 - \mathbf{m}_{b_1})^2}{\mathbf{s}_{b_1}^2}\right)\right\}.$$

a) Conjunta a posteriori.

$$\mathbf{p}_n(\mathbf{b}) \propto I_{[lb_0, upb_0]}(\mathbf{b}_0) \exp\left\{-\frac{1}{2\mathbf{s}^2 \mathbf{s}_{b_1}^2} (\mathbf{s}_{b_1}^2 (\mathbf{b}' Z_n Z_n \mathbf{b} - 2Y_n' Z_n \mathbf{b}) + \mathbf{s}^2 (\mathbf{b}_1^2 - 2\mathbf{b}_1 \mathbf{m}_{b_1}))\right\}.$$

b) Condicionales a posteriori.

$$\mathbf{p}_n(\mathbf{b}_0 | \mathbf{b}_1) \sim N^{[lb_0, upb_0]}(\bar{y}_n - \mathbf{b}_1 \bar{x}_n, \mathbf{s}^2/n),$$

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{b}_0) \sim \begin{cases} N\left(\frac{\mathbf{s}^2 \mathbf{m}_{b_1} - \mathbf{s}_{b_1}^2 (\mathbf{b}_0 \sum x_i - \sum x_i y_i)}{\mathbf{s}^2 + \mathbf{s}_{b_1}^2 \sum x_i^2}, \frac{\mathbf{s}^2 \mathbf{s}_{b_1}^2}{\mathbf{s}^2 + \mathbf{s}_{b_1}^2 \sum x_i^2}\right) & \text{si } \mathbf{b}_0 \in (lb_0, upb_0) \\ 0 & \text{en el resto.} \end{cases}$$

Éstas se pueden obtener de forma sencilla reagrupando el exponente de $\mathbf{p}_n(\mathbf{b})$ como

$$\frac{\mathbf{b}_0^2 n \mathbf{s}_{b_1}^2 + \mathbf{b}_1^2 (\mathbf{s}^2 + \mathbf{s}_{b_1}^2 \sum x_i^2) + 2\mathbf{b}_1 (\mathbf{s}_{b_1}^2 (\mathbf{b}_0 \sum x_i - \sum x_i y_i) - \mathbf{s}^2 \mathbf{m}_{b_1}) - 2\mathbf{s}_{b_1}^2 \mathbf{b}_0 \sum y_i}{2\mathbf{s}^2 \mathbf{s}_{b_1}^2}.$$

c) Marginales a posteriori.

$$\mathbf{p}_n(\mathbf{b}_0) \sim N^{[lb_0, upb_0]} \left(\frac{w_{1n} \hat{\mathbf{b}}_{0n}^{ls} + w_{2n} (\bar{y}_n - \mathbf{m}_{b_1} \bar{x}_n)}{w_{1n} + w_{2n}}, \frac{(\mathbf{s}_{b_1}^2 \sum x_i^2 + \mathbf{s}^2) \mathbf{s}^2}{w_{1n} + w_{2n}} \right),$$

$$\mathbf{p}_n(\mathbf{b}_1) \propto \left[F \left(\frac{upb_0 - E_{b_1}^*}{\sqrt{V_{b_1}^*}} \right) - F \left(\frac{lb_0 - E_{b_1}^*}{\sqrt{V_{b_1}^*}} \right) \right] \exp \left\{ \frac{1}{2} \left[\frac{(E_{b_1}^*)^2}{V_{b_1}^*} - \frac{\mathbf{b}_1^2 (\mathbf{s}^2 + \mathbf{s}_{b_1}^2 \sum x_i^2) - 2\mathbf{b}_1 (\mathbf{s}_{b_1}^2 \sum x_i y_i + \mathbf{s}^2 \mathbf{m}_{b_1})}{\mathbf{s}^2 \mathbf{s}_{b_1}^2} \right] \right\},$$

con $w_{1n} = n\mathbf{s}_{b_1}^2 \sum (x_i - \bar{x}_n)^2 = \mathbf{s}_{b_1}^2 |Z'_n Z_n|$, $w_{2n} = n\mathbf{s}^2$, $E_{b_1}^* = \bar{y}_n - \mathbf{b}_1 \bar{x}_n$ y $V_{b_1}^* = \mathbf{s}^2/n$.

d) Modas a posteriori.

\mathbf{b}_{0n}^{pmo} es el punto en $[lb_0, upb_0]$ que está más cerca de

$$\hat{\mathbf{b}}_{0n} = \frac{w_{1n} \hat{\mathbf{b}}_{0n}^{ls} + w_{2n} (\bar{y}_n - \mathbf{m}_{b_1} \bar{x}_n)}{w_{1n} + w_{2n}}.$$

Si $\hat{\mathbf{b}}_{0n} \in [lb_0, upb_0]$,

$$\mathbf{b}_{1n}^{pmo} = \frac{w_{1n} \hat{\mathbf{b}}_{1n}^{ls} + w_{2n} \mathbf{m}_{b_1}}{w_{1n} + w_{2n}}$$

mientras que si $\hat{\mathbf{b}}_{0n} \notin [lb_0, upb_0]$, \mathbf{b}_{0n}^{pmo} es el extremo del intervalo más próximo a $\hat{\mathbf{b}}_{0n}$, y si éste es lb_0

$$\mathbf{b}_{1n}^{pmo} = \frac{\mathbf{s}_{b_1}^2 (\sum x_i y_i - lb_0 \sum x_i) + \mathbf{s}^2 \mathbf{m}_{b_1}}{\mathbf{s}_{b_1}^2 \sum x_i^2 + \mathbf{s}^2}$$

(de forma análoga cuando $\mathbf{b}_{0n}^{pmo} = upb_0$).

Las modas a posteriori se obtienen maximizando el exponente de la densidad $\mathbf{p}_n(\mathbf{b})$ con la restricción de que $\mathbf{b}_0 \in (lb_0, upb_0)$. Ello lleva a resolver el sistema

$$\begin{cases} \mathbf{b}_0 n\mathbf{s}_{b_1}^2 + \mathbf{b}_1 \mathbf{s}_{b_1}^2 \sum x_i - \mathbf{s}_{b_1}^2 \sum y_i = 0 \\ \mathbf{b}_0 \mathbf{s}_{b_1}^2 \sum x_i + \mathbf{b}_1 (\mathbf{s}^2 + \mathbf{s}_{b_1}^2 \sum x_i^2) - \mathbf{s}_{b_1}^2 \sum x_i y_i - \mathbf{s}^2 \mathbf{m}_{b_1} = 0. \end{cases}$$

e) Varianzas y Covarianzas de las modas a posteriori.

Si $\mathbf{b}_{0n}^{pmo} \in (lb_0, upb_0)$, la matriz de varianzas covarianzas es

$$Var_Cov_n(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo} | \mathbf{b}_{0n}^{pmo} \in (lb_0, upb_0)) = \frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \begin{pmatrix} w_{1n}^2 (Z_n' Z_n)^{-1} + (2w_{1n}w_{2n} + w_{2n}^2) \begin{pmatrix} 1/n & 0 \\ 0 & 0 \end{pmatrix} \\ \end{pmatrix}.$$

Desarrollando se tiene que

$$V_n(\mathbf{b}_{0n}^{pmo} | \mathbf{b}_{0n}^{pmo} \in (lb_0, upb_0)) = \frac{1}{(w_{1n} + w_{2n})^2} V_n(w_{1n} \hat{\mathbf{b}}_{0n}^{ls} + w_{2n}(\bar{y}_n - \bar{x}_n \mathbf{m}_{b_1})) = \frac{1}{(w_{1n} + w_{2n})^2} \left(w_{1n}^2 \frac{\mathbf{s}^2 \sum x_i^2}{n \sum (x_i - \bar{x}_n)^2} + w_{2n}^2 \frac{\mathbf{s}^2}{n} + 2w_{1n}w_{2n} Cov_n(\hat{\mathbf{b}}_{0n}^{ls}, \bar{y}_n) \right) = \frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \left(w_{1n}^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x}_n)^2} + \frac{w_{2n}^2 + 2w_{1n}w_{2n}}{n} \right)$$

$$V_n(\mathbf{b}_{1n}^{pmo} | \mathbf{b}_{0n}^{pmo} \in (lb_0, upb_0)) = \frac{1}{(w_{1n} + w_{2n})^2} V_n(w_{1n} \hat{\mathbf{b}}_{1n}^{ls}) = \frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \frac{w_{1n}^2}{\sum (x_i - \bar{x}_n)^2},$$

$$Cov_n(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo} | \mathbf{b}_{0n}^{pmo} \in (lb_0, upb_0)) = \frac{1}{(w_{1n} + w_{2n})^2} \left(w_{1n}^2 Cov_n(\hat{\mathbf{b}}_{0n}^{ls}, \hat{\mathbf{b}}_{1n}^{ls}) + w_{1n} \mathbf{s}^2 Cov_n(\sum y_i, \hat{\mathbf{b}}_{1n}^{ls}) \right) = \frac{\mathbf{s}^2}{(w_{1n} + w_{2n})^2} \begin{pmatrix} -w_{1n}^2 \sum x_i \\ n \sum (x_i - \bar{x}_n)^2 \end{pmatrix}.$$

Si $\mathbf{b}_{0n}^{pmo} \notin (lb_0, upb_0)$,

$$V_n(\mathbf{b}_{1n}^{pmo}) = \frac{\mathbf{s}_{b_1}^4 \mathbf{s}^2 \sum x_i^2}{(\mathbf{s}^2 + \mathbf{s}_{b_1}^2 \sum x_i^2)^2}.$$

f) Varianza de estimación de predicción, $f_n^{pm}(x_0; \mathbf{b}) = \mathbf{b}_{0n}^{pmo} + \mathbf{b}_{1n}^{pmo} x_0$.

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{b}_{0n}^{pmo} \in (lb_0, upb_0)) = \mathbf{s}^2 \left(\frac{1}{n} + \frac{w_{1n}^2}{(w_{1n} + w_{2n})^2} \frac{(x_0 - \bar{x}_n)^2}{\sum (x_i - \bar{x}_n)^2} \right),$$

con $w_{1n} = n \mathbf{s}_{b_1}^2 \sum (x_i - \bar{x}_n)^2 = \mathbf{s}_{b_1}^2 |Z_n' Z_n|$, $w_{2n} = n \mathbf{s}^2$ y

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{b}_{0n}^{pmo} = lb_0) = \frac{x_0^2 \mathbf{s}_{b_1}^4 \mathbf{s}^2 \sum x_i^2}{(\mathbf{s}^2 + \mathbf{s}_{b_1}^2 \sum x_i^2)^2}.$$

De forma análoga se obtienen para $\mathbf{b}_{0n}^{pmo} = upb_0$.

4. PRIORI UNIFORME SOBRE b_0 y b_1 INDEPENDIENTES.

Consideremos en este caso que $\mathbf{p}_0(\mathbf{b})$ indica que $\mathbf{b}_0 \sim U[lb_0, upb_0]$ y $\mathbf{b}_1 \sim U[lb_1, upb_1]$ independientes. La densidad a priori es

$$\mathbf{p}_0(\mathbf{b}) \propto I_{[lb_0, upb_0]}(\mathbf{b}_0) I_{[lb_1, upb_1]}(\mathbf{b}_1).$$

La distribución conjunta de Y_n y \mathbf{b} queda como sigue

$$h(Y_n, \mathbf{b} | X_n, \mathbf{p}_0) \propto I_{[lb_0, upb_0]}(\mathbf{b}_0) I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{-\frac{1}{2\mathbf{S}^2} (Y_n - Z_n \mathbf{b})'(Y_n - Z_n \mathbf{b})\right\}.$$

a) Conjunta a posteriori.

$$\mathbf{p}_n(\mathbf{b}) \propto I_{[lb_0, upb_0]}(\mathbf{b}_0) I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{-\frac{1}{2\mathbf{S}^2} (\mathbf{b}' Z_n' Z_n \mathbf{b} - 2Y_n' Z_n \mathbf{b})\right\}.$$

b) Condicionales a posteriori.

$$\mathbf{p}_n(\mathbf{b}_0 | \mathbf{b}_1) \sim \begin{cases} N^{[lb_0, upb_0]}(\bar{y}_n - \mathbf{b}_1 \bar{x}_n, \mathbf{S}^2/n) & \text{si } \mathbf{b}_1 \in [lb_1, upb_1] \\ 0 & \text{en el resto.} \end{cases}$$

Cuando $\sum x_i^2 \neq 0$,

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{b}_0) \sim \begin{cases} N^{[lb_1, upb_1]}((\sum x_i y_i - \mathbf{b}_0 \sum x_i) / \sum x_i^2, \mathbf{S}^2 / \sum x_i^2) & \mathbf{b}_0 \in [lb_0, upb_0] \\ 0 & \text{en el resto.} \end{cases}$$

Mientras si $\sum x_i^2 = 0$,

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{b}_0) \sim \begin{cases} U[lb_1, upb_1] & \text{si } \mathbf{b}_0 \in [lb_0, upb_0] \\ 0 & \text{en el resto.} \end{cases}$$

Estas expresiones se obtiene teniendo en cuenta que el exponente de $\mathbf{p}_n(\mathbf{b})$ se reescribe como

$$-\frac{1}{2\mathbf{S}^2} (n\mathbf{b}_0^2 + \mathbf{b}_1^2 \sum x_i^2 - 2\mathbf{b}_1 (\sum x_i y_i - \mathbf{b}_0 \sum x_i) - 2\mathbf{b}_0 \sum y_i).$$

c) Marginales a posteriori.

Cuando $\sum x_i^2 \neq 0$,

$$\mathbf{p}_n(\mathbf{b}_0) \propto I_{[lb_0, upb_0]}(\mathbf{b}_0) \left[\mathbf{F} \left(\frac{upb_0 - E_{\mathbf{b}_0}^*}{\sqrt{V_{\mathbf{b}_0}^*}} \right) - \mathbf{F} \left(\frac{lb_0 - E_{\mathbf{b}_0}^*}{\sqrt{V_{\mathbf{b}_0}^*}} \right) \right] \exp \left\{ \frac{1}{2} \left[\frac{(E_{\mathbf{b}_0}^*)^2}{V_{\mathbf{b}_0}^*} - \frac{n\mathbf{b}_0^2 - 2\mathbf{b}_0 \sum y_i}{\mathbf{s}^2} \right] \right\},$$

$$\text{con } E_{\mathbf{b}_0}^* = \frac{\sum x_i y_i - \mathbf{b}_0 \sum x_i}{\sum x_i^2} \text{ y } V_{\mathbf{b}_0}^* = \mathbf{s}^2 / \sum x_i^2.$$

Mientras que si $\sum x_i^2 = 0$,

$$\mathbf{p}_n(\mathbf{b}_0) \propto I_{[lb_0, upb_0]} \exp \left\{ -\frac{n(\mathbf{b}_0)^2 - 2\mathbf{b}_0 \sum y_i}{2\mathbf{s}^2} \right\},$$

$$\mathbf{p}_n(\mathbf{b}_1) \propto I_{[lb_1, upb_1]}(\mathbf{b}_1) \left[\mathbf{F} \left(\frac{upb_0 - E_{\mathbf{b}_1}^*}{\sqrt{V_{\mathbf{b}_1}^*}} \right) - \mathbf{F} \left(\frac{lb_0 - E_{\mathbf{b}_1}^*}{\sqrt{V_{\mathbf{b}_1}^*}} \right) \right] \exp \left\{ \frac{1}{2} \left[\frac{(E_{\mathbf{b}_1}^*)^2}{V_{\mathbf{b}_1}^*} - \frac{\mathbf{b}_1^2 \sum x_i^2 - 2\mathbf{b}_1 \sum x_i y_i}{\mathbf{s}^2} \right] \right\},$$

$$\text{con } E_{\mathbf{b}_1}^* = \bar{y}_n - \mathbf{b}_1 \bar{x}_n \text{ y } V_{\mathbf{b}_1}^* = \mathbf{s}^2 / n.$$

d) Modas a posteriori.

Si $Z_n' Z_n$ es invertible:

- Si $\hat{\mathbf{b}}_{0n}^{ls} \in [lb_0, upb_0]$ y $\hat{\mathbf{b}}_{1n}^{ls} \in [lb_1, upb_1]$, $(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo}) = (\hat{\mathbf{b}}_{0n}^{ls}, \hat{\mathbf{b}}_{1n}^{ls})$.
- Si $\hat{\mathbf{b}}_{0n}^{ls} \in [lb_0, upb_0]$ y $\hat{\mathbf{b}}_{1n}^{ls} \notin [lb_1, upb_1]$ entonces \mathbf{b}_{1n}^{pmo} es el extremo del intervalo más próximo a $\hat{\mathbf{b}}_{1n}^{ls}$ y \mathbf{b}_{0n}^{pmo} es el punto de $[lb_0, upb_0]$ más próximo a $\bar{y}_n - \bar{x}_n \mathbf{b}_{1n}^{pmo}$.
- Si $\hat{\mathbf{b}}_{0n}^{ls} \notin [lb_0, upb_0]$ y $\hat{\mathbf{b}}_{1n}^{ls} \in [lb_1, upb_1]$, \mathbf{b}_{0n}^{pmo} es el extremo del intervalo más próximo a $\hat{\mathbf{b}}_{0n}^{ls}$ y \mathbf{b}_{1n}^{pmo} es el punto de $[lb_1, upb_1]$ más próximo a $(\sum x_i y_i - \hat{\mathbf{b}}_{0n}^{pmo} \sum x_i) / \sum x_i^2$.
- Si $\hat{\mathbf{b}}_{0n}^{ls} \notin [lb_0, upb_0]$ y $\hat{\mathbf{b}}_{1n}^{ls} \notin [lb_1, upb_1]$, $(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo})$ será el punto que minimiza $g(\mathbf{b}_0, \mathbf{b}_1) = \mathbf{b}' Z_n' Z_n \mathbf{b} - 2Y_n' Z_n \mathbf{b}$ entre los dos $(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo})$ obtenidos en b) y c).

Si $Z'_n Z_n$ no es invertible y la recta $\mathbf{b}_0 + \mathbf{b}_1 x_1 = \bar{y}_n$ tiene intersección con el recinto $[lb_0, upb_0] \times [lb_1, upb_1]$, $(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo})$ es cualquier punto de dicha recta que también esté en el recinto. Si la recta no tiene intersección con el recinto, $(\mathbf{b}_{0n}^{pmo}, \mathbf{b}_{1n}^{pmo})$ será el punto esquina del recinto más próximo a dicha recta.

Las modas a posteriori se obtendrán maximizando el exponente de la densidad $\mathbf{p}_n(\mathbf{b})$ con las restricciones de que $\mathbf{b}_0 \in [lb_0, upb_0]$ y $\mathbf{b}_1 \in [lb_1, upb_1]$. Ello lleva a resolver el sistema

$$\begin{cases} n\mathbf{b}_0 + \mathbf{b}_1 \sum x_i - \sum y_i = 0 \\ \mathbf{b}_0 \sum x_i + \mathbf{b}_1 \sum x_i^2 - \sum x_i y_i = 0. \end{cases}$$

e) Varianza de estimación de predicción, $f_n^{pm}(x_0; \mathbf{b}) = \mathbf{b}_{0n}^{pmo} + \mathbf{b}_{1n}^{pmo} x_0$.

Cuando $Z'_n Z_n$ es invertible,

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{b}_{0n}^{pmo} \in (lb_0, upb_0), \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) = \mathbf{s}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x}_n)^2}{\sum (x_i - \bar{x}_n)^2} \right),$$

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{b}_{0n}^{pmo} \in (lb_0, upb_0), \mathbf{b}_{1n}^{pmo} = lb_1) = \frac{\mathbf{s}^2}{n},$$

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{b}_{0n}^{pmo} = lb_0, \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) = \frac{\mathbf{s}^2 x_0^2}{\sum x_i^2}.$$

De forma análoga se obtiene cuando $\mathbf{b}_{0n}^{pmo} = upb_0$ y $\mathbf{b}_{1n}^{pmo} = upb_1$.

En el caso de que $Z'_n Z_n$ sea no invertible no se especifica la varianza de predicción ya que las modas a posteriori no quedan definidas de forma unívoca.

5. ESTIMADORES MÍNIMOS CUADRADOS.

A continuación se presentan algunos resultados para los estimadores mínimos cuadrados que son de gran utilidad en el desarrollo de este apéndice.

$$\hat{\mathbf{b}}_{0n}^{ls} = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum (x_i - \bar{x}_n)^2}, \quad \hat{\mathbf{b}}_{1n}^{ls} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum (x_i - \bar{x}_n)^2}.$$

$$Var_Cov_n(\hat{\mathbf{b}}_{0n}^{ls}, \hat{\mathbf{b}}_{1n}^{ls}) = \mathbf{s}^2 (Z_n' Z_n)^{-1} = \frac{\mathbf{s}^2}{n \sum (x_i - \bar{x}_n)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix},$$

$$Cov_n(\hat{\mathbf{b}}_{1n}^{ls}, \sum x_i y_i) = Cov\left(\frac{\sum y_i (x_i - \bar{x}_n)}{\sum (x_i - \bar{x}_n)^2}, \sum x_i y_i\right) = \mathbf{s}^2 \frac{\sum (x_i - \bar{x}_n) x_i}{\sum (x_i - \bar{x}_n)^2} = \mathbf{s}^2,$$

$$Cov_n(\hat{\mathbf{b}}_{1n}^{ls}, \sum y_i) = Cov\left(\frac{\sum y_j (n x_j - \sum x_i)}{n \sum (x_i - \bar{x}_n)^2}, \sum y_i\right) = \mathbf{s}^2 \frac{\sum (n x_j - \sum x_i)}{n \sum (x_i - \bar{x}_n)^2} = 0,$$

$$Cov_n(\hat{\mathbf{b}}_{0n}^{ls}, \sum x_i y_i) = Cov_n(\sum y_i / n, \sum x_i y_i) - \bar{x}_n Cov_n(\hat{\mathbf{b}}_{1n}^{ls}, \sum x_i y_i) = \mathbf{s}^2 \bar{x}_n - \bar{x}_n \mathbf{s}^2 = 0,$$

$$Cov_n(\hat{\mathbf{b}}_{0n}^{ls}, \sum y_i) = Cov_n(\bar{y}_n, \sum y_i) = \mathbf{s}^2,$$

$$V_n(\hat{\mathbf{b}}_{0n}^{ls} + \hat{\mathbf{b}}_{1n}^{ls} x_0) = \mathbf{s}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x}_n)^2}{\sum (x_i - \bar{x}_n)^2} \right).$$

APENDICE II

RESULTADOS DISTRIBUCIONALES: MODELO DE REGRESIÓN SIMPLE NORMAL PARAMETRIZADO EN TÉRMINOS DE b_1 y q .

En el Apéndice I se trató el modelo de regresión simple normal parametrizado por $E[y_n | x_n, \mathbf{b}] = \mathbf{b}_0 + \mathbf{b}_1 x_n$. En ese caso, la raíz queda expresada como el cociente $\mathbf{q} = -\mathbf{b}_0/\mathbf{b}_1$, lo que lleva, en el caso de que la distribución sobre \mathbf{b}_0 tenga soporte infinito o la de \mathbf{b}_1 contenga el cero, a que el soporte de la distribución de \mathbf{q} sea igual a R . A menudo esto no refleja la incertidumbre sobre dicho parámetro \mathbf{q} de forma adecuada y es corriente que a los científicos les sea más fácil ofrecer información sobre la localización de la raíz y sobre la pendiente de la curva cerca de dicha raíz. Para cubrir esta posibilidad se considera el modelo parametrizado de la forma $E[y_n | x_n, \mathbf{q}, \mathbf{b}_1] = T + \mathbf{b}_1(x_n - \mathbf{q})$ con diferentes distribuciones a priori sobre los parámetros.

Asumamos que

$$y_n | x_n, \mathbf{b}_1, \mathbf{q} \sim N(T + \mathbf{b}_1(x_n - \mathbf{q}), \mathbf{s}^2)$$

con las y_n independientes o, en forma matricial,

$$Y_n | X_n, \mathbf{b}_1, \mathbf{q} \sim N(T + \mathbf{b}_1 Z_n \mathbf{q}_1, \mathbf{s}^2 I_n),$$

con Z_n la matriz de $n \times 2$ cuya j -ésima fila es el vector $(1, x_j)$, $Y_n' = (y_1, \dots, y_n)$, $\mathbf{q}_1' = (-\mathbf{q}, 1)$ e I_n la matriz identidad $n \times n$, con \mathbf{s} conocida y T un valor objetivo también conocido. Sin pérdida de generalidad, supondremos $T = 0$, restando el valor T de y_i . Como en el apéndice anterior a continuación se presentan las distribuciones a posteriori de los parámetros, las

condicionales a posteriori de un parámetro respecto al otro y las marginales a posteriori, todo para cuatro distribuciones a priori formadas al combinar la distribución normal y la uniforme sobre \mathbf{q} y \mathbf{b}_1 , bajo la suposición inicial de independencia. También se indicará la moda a posteriori de los parámetros para las dos prioris para las que $\mathbf{q} \sim U[rtl, rtu]$. No se han calculado las modas a posteriori para los casos en los que \mathbf{q} fuese normal, porque no es posible expresarlas en forma cerrada, ya que para ello es necesario resolver un sistema no lineal de difícil solución. En estos casos, para el cálculo de esperanzas, modas y varianzas a posteriori, se deberán recurrir a métodos de simulación, tales como son MonteCarlo, Rejection Sampling y Gibbs. A lo largo del apéndice la notación empleada será la que se especifica en la Sección 1.7 del primer capítulo.

Para la parametrización de este apéndice, estimaremos $E[y|x_0] = f(x_0; \mathbf{b})$ a través de $f_n^{pm}(x_0; \mathbf{b}) = \mathbf{b}_{ln}^{pmo}(x_0 - \mathbf{q}_n^{pmo})$ donde $(\mathbf{q}_n^{pmo}, \mathbf{b}_{ln}^{pmo})$ es la moda de la distribución a posteriori $\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1)$. A pesar de que \mathbf{q}_n^{pmo} no es el mejor estimador para \mathbf{q} , ya que prefiere a los valores de rtl ó rtu sobre puntos del interior de $[rtl, rtu]$, tiene la ventaja de que tanto \mathbf{q}_n^{pmo} como $f_n^{pm}(x_0; \mathbf{b})$ y la desviación estándar de éste último pueden evaluarse de forma cerrada. Esto hace factible su utilización en los ejemplos de esta memoria y permite apreciar de una manera sencilla la generalidad de los diseños propuestos en los distintos capítulos al ilustrar su uso sobre distintas versiones de *certainty equivalence*.

1. PRIORI UNIFORME EN q Y NORMAL EN b_1 .

Consideremos que la distribución a priori sobre los parámetros $\mathbf{p}_0(\mathbf{q}, \mathbf{b}_1)$ es tal que $\mathbf{q} \sim U[rtl, rtu]$, $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1}, \mathbf{s}_{b_1}^2)$ y \mathbf{q} y \mathbf{b}_1 independientes. La densidad a priori es

$$\mathbf{p}_0(\mathbf{q}, \mathbf{b}_1) \propto I_{[rtl, rtu]}(\mathbf{q}) \exp\left\{-\frac{1}{2}\left(\frac{\mathbf{b}_1 - \mathbf{m}_{b_1}}{\mathbf{s}_{b_1}}\right)^2\right\},$$

donde I representa la función indicatriz. La distribución conjunta de Y_n , \mathbf{q} y \mathbf{b}_1 queda como

$$h(Y_n, \mathbf{q}, \mathbf{b}_1 | X_n, \mathbf{p}_0) \propto I_{[rtl, rtu]}(\mathbf{q}) \exp\left\{-\frac{1}{2}\left(\frac{(Y_n - \mathbf{b}_1 Z_n \mathbf{q}_1)'(Y_n - \mathbf{b}_1 Z_n \mathbf{q}_1)}{\mathbf{s}^2} + \frac{(\mathbf{b}_1 - \mathbf{m}_{b_1})^2}{\mathbf{s}_{b_1}^2}\right)\right\}.$$

a) Conjunta a posteriori.

$$\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1) \propto I_{[rtl, rtu]}(\mathbf{q}) \exp\left\{-\frac{1}{2\mathbf{s}^2 \mathbf{s}_{b_1}^2} \left(\mathbf{b}_1^2 (\mathbf{s}_{b_1}^2 \mathbf{q}_1' Z_n' Z_n \mathbf{q}_1 + \mathbf{s}^2) - 2\mathbf{b}_1 (\mathbf{s}_{b_1}^2 Y_n' Z_n \mathbf{q}_1 + \mathbf{s}^2 \mathbf{m}_{b_1})\right)\right\}.$$

b) Condicionales a posteriori.

$$\mathbf{p}_n(\mathbf{q} | \mathbf{b}_1) \sim \begin{cases} N^{[rtl, rtu]}(\bar{x}_n - \bar{y}_n / \mathbf{b}_1, \mathbf{s}^2 / (n\mathbf{b}_1^2)) & \text{si } \mathbf{b}_1 \neq 0 \\ U[rtl, rtu] & \text{si } \mathbf{b}_1 = 0 \end{cases}$$

y

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{q}) \sim \begin{cases} N\left(\frac{\mathbf{s}_{b_1}^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 \mathbf{m}_{b_1}}{\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2}, \frac{\mathbf{s}_{b_1}^2 \mathbf{s}^2}{\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2}\right) & \text{si } \mathbf{q} \in [rtl, rtu] \\ 0 & \text{en el resto.} \end{cases}$$

Éstas se obtienen desarrollando la parte del exponente de $\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1)$ que depende de \mathbf{q} y \mathbf{b}_1 :

$$-\frac{\mathbf{b}_1^2 (\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2) - 2\mathbf{b}_1 (\mathbf{s}_{b_1}^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 \mathbf{m}_{b_1})}{2\mathbf{s}^2 \mathbf{s}_{b_1}^2}.$$

c) Marginales a posteriori.

$$p_n(\mathbf{q}) \propto I_{[rtl,rtu]}(\mathbf{q}) \frac{1}{\sqrt{\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2}} \exp \left\{ \frac{1}{2} \frac{(\mathbf{s}_{b_1}^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 \mathbf{m}_{b_1})^2}{\mathbf{s}^2 \mathbf{s}_{b_1}^2 (\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2)} \right\}$$

y

$$p_n(\mathbf{b}_1) \propto \sqrt{V_{b_1}^*} \left[F \left(\frac{rtu - E_{b_1}^*}{\sqrt{V_{b_1}^*}} \right) - F \left(\frac{rtl - E_{b_1}^*}{\sqrt{V_{b_1}^*}} \right) \right] \exp \left\{ \frac{1}{2} \left[\frac{(E_{b_1}^*)^2}{V_{b_1}^*} - \frac{\mathbf{b}_1^2 (\mathbf{s}_{b_1}^2 \sum x_i^2 + \mathbf{s}^2) - 2 \mathbf{b}_1 (\mathbf{s}_{b_1}^2 \sum x_i y_i + \mathbf{s}^2 \mathbf{m}_{b_1})}{\mathbf{s}^2} \right] \right\},$$

con $E_{b_1}^* = \bar{x}_n - \bar{y}_n / \mathbf{b}_1$, $V_{b_1}^* = \mathbf{s}^2 / (n \mathbf{b}_1^2)$.

La derivada de la densidad marginal a posteriori de θ igualada a cero resulta el polinomio de tercer grado

$$\begin{aligned} & \mathbf{q}^3 (\mathbf{s}_{b_1}^4 \mathbf{s}^2 n^2) + \mathbf{q}^2 (-3 n \mathbf{s}_{b_1}^4 \mathbf{s}^2 \sum x_i - n \text{cte}2 \mathbf{s}_{b_1}^2 \sum y_i + \mathbf{s}_{b_1}^4 (\sum y_i)^2 \sum x_i) \\ & + \mathbf{q} (n \text{cte}2^2 + n \mathbf{s}_{b_1}^2 \mathbf{s}^2 \text{cte}1 + 2 \mathbf{s}_{b_1}^4 \mathbf{s}^2 (\sum x_i)^2 - \text{cte}1 \mathbf{s}_{b_1}^2 (\sum y_i)^2) \\ & - \sum x_i (\text{cte}2^2 + \mathbf{s}_{b_1}^2 \mathbf{s}^2 \text{cte}1) + \text{cte}1 \text{cte}2 \sum y_i = 0 \end{aligned}$$

con $\text{cte}1 = \mathbf{s}_{b_1}^2 \sum x_i^2 + \mathbf{s}^2$ y $\text{cte}2 = \mathbf{s}_{b_1}^2 \sum x_i y_i + \mathbf{s}^2 \mathbf{m}_{b_1}$, lo que indica que a lo sumo la distribución es bimodal.

d) Modas a posteriori.

\mathbf{q}_n^{pmo} es el punto de $[rtl,rtu]$ más cercano a

$$\hat{\mathbf{q}}_n = - \frac{\mathbf{s}_{b_1}^2 (\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i) + \mathbf{s}^2 (\sum y_i - \mathbf{m}_{b_1} \sum x_i)}{n \mathbf{s}_{b_1}^2 \sum (x_i - \bar{x}_n)(y_i - \bar{y}_n) + n \mathbf{s}^2 \mathbf{m}_{b_1}} = - \frac{w_{1n} \hat{\mathbf{b}}_{0n}^{ls} + w_{2n} (\bar{y}_n - \mathbf{m}_{b_1} \bar{x}_n)}{w_{1n} \hat{\mathbf{b}}_{1n}^{ls} + w_{2n} \mathbf{m}_{b_1}},$$

con $w_{1n} = n \mathbf{s}_{b_1}^2 \sum (x_i - \bar{x}_n)^2 = \mathbf{s}_{b_1}^2 |Z_n' Z_n|$ y $w_{2n} = n \mathbf{s}^2$.

Si $\hat{\mathbf{q}}_n \in [rtl,rtu]$,

$$\mathbf{b}_{1n}^{pmo} = \frac{n \mathbf{s}_{b_1}^2 \sum (x_i - \bar{x}_n)(y_i - \bar{y}_n) + n \mathbf{s}^2 \mathbf{m}_{b_1}}{n \mathbf{s}_{b_1}^2 \sum (x_i - \bar{x}_n)^2 + n \mathbf{s}^2} = \frac{w_{1n} \hat{\mathbf{b}}_{1n}^{ls} + w_{2n} \mathbf{m}_{b_1}}{w_{1n} + w_{2n}},$$

mientras que si $\hat{q}_n \notin [rtl, rtu]$, q_n^{pmo} es el extremo del intervalo más próximo a \hat{q}_n , y si éste es rtl ,

$$b_{1n}^{pmo} = \frac{s_{b_1}^2 \sum (x_i - rtl)y_i + s^2 m_{b_1}}{s_{b_1}^2 \sum (x_i - rtl)^2 + s^2}$$

(de forma análoga cuando $q_n^{pmo} = rtu$).

Estas modas a posteriori se obtendrán de maximizar el exponente de la densidad $p_n(\mathbf{q}, \mathbf{b}_1)$ con la restricción de que $\mathbf{q} \in [rtl, rtu]$. Ello lleva a resolver el sistema

$$\begin{cases} b_1 (s_{b_1}^2 \sum (x_i - \mathbf{q})^2 + s^2) - (s_{b_1}^2 (\sum x_i y_i - \mathbf{q} \sum y_i) + s^2 m_{b_1}) = 0 \\ b_1^2 (n\mathbf{q} - \sum x_i) + b_1 \sum y_i = 0. \end{cases}$$

e) Varianza de estimación de predicción, $f_n^{pm}(x_0; \mathbf{b}) = b_{1n}^{pmo}(x_0 - q_n^{pmo})$.

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | q_n^{pmo} \in (rtl, rtu)) = s^2 \left(\frac{1}{n} + \frac{w_{1n}^2}{(w_{1n} + w_{2n})^2} \frac{(x_0 - \bar{x}_n)^2}{\sum (x_i - \bar{x}_n)^2} \right),$$

con $w_{1n} = n s_{b_1}^2 \sum (x_i - \bar{x}_n)^2 = s_{b_1}^2 |Z_n' Z_n|$ y $w_{2n} = n s^2$,

$$\begin{aligned} V_n(f_n^{pm}(x_0; \mathbf{b}) | q_n^{pmo} = rtl) &= (x_0 - rtl)^2 V_n(b_{1n}^{pmo} | q_n^{pmo} = rtl) = \\ &= (x_0 - rtl)^2 \frac{s_{b_1}^4 s^2 \sum (x_i - rtl)^2}{(s_{b_1}^2 \sum (x_i - rtl)^2 + s^2)^2}. \end{aligned}$$

De forma similar se obtiene para $q_n^{pmo} = rtu$.

2. PRIORI UNIFORME EN \mathbf{q} Y EN \mathbf{b}_1 .

Consideremos que la distribución a priori sobre los parámetros $\mathbf{p}_0(\mathbf{q}, \mathbf{b}_1)$ es tal que $\mathbf{q} \sim U[rtl, rtu]$ y $\mathbf{b}_1 \sim U[lb_1, upb_1]$ independientes. La densidad a priori es

$$\mathbf{p}_0(\mathbf{q}, \mathbf{b}_1) \propto I_{[rtl, rtu]}(\mathbf{q}) I_{[lb_1, upb_1]}(\mathbf{b}_1).$$

La distribución conjunta de Y_n , \mathbf{q} y \mathbf{b}_1 queda como sigue

$$h(Y_n, \mathbf{q}, \mathbf{b}_1 | X_n, \mathbf{p}_0) \propto I_{[rtl, rtu]}(\mathbf{q}) I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{-\frac{1}{2} \frac{(Y_n - \mathbf{b}_1 Z_n \mathbf{q}_1)'(Y_n - \mathbf{b}_1 Z_n \mathbf{q}_1)}{\mathbf{s}^2}\right\}.$$

a) Conjunta a posteriori.

$$\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1) \propto I_{[rtl, rtu]}(\mathbf{q}) I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{-\frac{1}{2\mathbf{s}^2} (\mathbf{b}_1^2 \mathbf{q}_1' Z_n' Z_n \mathbf{q}_1 - 2\mathbf{b}_1 Y_n' Z_n \mathbf{q}_1)\right\}.$$

b) Condicionales a posteriori.

$$\mathbf{p}_n(\mathbf{q} | \mathbf{b}_1) \sim \begin{cases} N^{[rtl, rtu]}(\bar{x}_n - \bar{y}_n / \mathbf{b}_1, \mathbf{s}^2 / (n\mathbf{b}_1^2)) & \text{si } \mathbf{b}_1 \in [lb_1, upb_1] \\ 0 & \text{en el resto} \end{cases}$$

y

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{q}) \propto \begin{cases} N^{[lb_1, upb_1]} \left(\frac{\sum y_i (x_i - \mathbf{q})}{\sum (x_i - \mathbf{q})^2}, \frac{\mathbf{s}^2}{\sum (x_i - \mathbf{q})^2} \right) & \text{si } \mathbf{q} \in [rtl, rtu] \\ 0 & \text{en el resto.} \end{cases}$$

Éstas se obtienen desarrollando la parte del exponente de $\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1)$ que depende de \mathbf{q} y \mathbf{b}_1 :

$$-\frac{1}{2\mathbf{s}^2} (\mathbf{b}_1^2 \sum (x_i - \mathbf{q})^2 - 2\mathbf{b}_1 \sum y_i (x_i - \mathbf{q})).$$

c) Marginales a posteriori.

$$\mathbf{p}_n(\mathbf{q}) \propto I_{[rtl,rtu]}(\mathbf{q}) \sqrt{V_q^*} \left[\mathbf{F} \left(\frac{upb_1 - E_q^*}{\sqrt{V_q^*}} \right) - \mathbf{F} \left(\frac{lb_1 - E_q^*}{\sqrt{V_q^*}} \right) \right] \exp \left\{ \frac{(E_q^*)^2}{2V_q^*} \right\},$$

con $E_q^* = \frac{\sum y_i(x_i - \mathbf{q})}{\sum (x_i - \mathbf{q})^2}$ y $V_q^* = \mathbf{s}^2 / \sum (x_i - \mathbf{q})^2$.

$$\mathbf{p}_n(\mathbf{b}_1) \propto I_{[lb_1,upb_1]} \sqrt{V_{b_1}^*} \left[\mathbf{F} \left(\frac{rtu - E_{b_1}^*}{\sqrt{V_{b_1}^*}} \right) - \mathbf{F} \left(\frac{rtl - E_{b_1}^*}{\sqrt{V_{b_1}^*}} \right) \right] \exp \left\{ -\frac{1}{2\mathbf{s}^2} (\mathbf{b}_1^2 \sum x_i(x_i - \bar{x}_n) - 2\mathbf{b}_1 \sum x_i(y_i - \bar{y}_n)) \right\},$$

con $E_{b_1}^* = \bar{x} - \bar{y}_n / \mathbf{b}_1$ y $V_{b_1}^* = \mathbf{s}^2 / n\mathbf{b}_1^2$.

d) Modas a posteriori.

Si $Z'_n Z_n$ es invertible y $\hat{\mathbf{q}}_n = -\frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i y_i - \sum x_i \sum y_i}$ se tiene que:

- a) Si $\hat{\mathbf{q}}_n \in [rtl,rtu]$ y $\hat{\mathbf{b}}_{1n}^{ls} \in [lb_1,upb_1]$, $(\mathbf{q}_n^{pmo}, \mathbf{b}_{1n}^{pmo}) = (\hat{\mathbf{q}}_n, \hat{\mathbf{b}}_{1n}^{ls})$.
- b) Si $\hat{\mathbf{q}}_n \in [rtl,rtu]$ y $\hat{\mathbf{b}}_{1n}^{ls} \notin [lb_1,upb_1]$, \mathbf{b}_{1n}^{pmo} es el extremo del intervalo $[lb_1,upb_1]$ más próximo a $\hat{\mathbf{b}}_{1n}^{ls}$ y \mathbf{q}_n^{pmo} es el punto de $[rtl,rtu]$ más próximo a $\bar{x}_n - \bar{y}_n / \mathbf{b}_{1n}^{pmo}$.
- c) Si $\hat{\mathbf{q}}_n \notin [rtl,rtu]$ y $\hat{\mathbf{b}}_{1n}^{ls} \in [lb_1,upb_1]$, \mathbf{q}_n^{pmo} es el extremo del intervalo $[rtl,rtu]$ más próximo a $\hat{\mathbf{q}}_n$ y \mathbf{b}_{1n}^{pmo} es el punto de $[lb_1,upb_1]$ más próximo a $\sum y_i(x_i - \mathbf{q}_n^{pmo}) / \sum (x_i - \mathbf{q}_n^{pmo})^2$.
- d) Si $\hat{\mathbf{q}}_n \notin [rtl,rtu]$ y $\hat{\mathbf{b}}_{1n}^{ls} \notin [lb_1,upb_1]$, $(\mathbf{q}_n^{pmo}, \mathbf{b}_{1n}^{pmo})$ es el punto que minimiza $g(\mathbf{q}, \mathbf{b}_1) = \mathbf{b}_1^2 \sum (x_i - \mathbf{q}) - 2\mathbf{b}_1 \sum y_i(x_i - \mathbf{q})$ entre los puntos elegidos como en b) y c).

Mientras si $Z'_n Z_n$ no es invertible y la hipérbola $\mathbf{b}_1(\mathbf{q} - x_i) = -\bar{y}_n$ tiene intersección con el recinto $[rtl,rtu] \times [lb_1,upb_1]$, $(\mathbf{q}_n^{pmo}, \mathbf{b}_{1n}^{pmo})$ podrá ser cualquier punto de dicha hipérbola que

esté dentro del recinto. En caso contrario, $(\mathbf{q}_n^{pmo}, \mathbf{b}_{1n}^{pmo})$ será el punto esquina del recinto más próximo a dicha hipérbola.

Estas modas a posteriori se obtienen maximizando el exponente de la densidad $p_n(\mathbf{q}, \mathbf{b}_1)$ con la restricción de que $\mathbf{q} \in (rtl, rtu)$ y $\mathbf{b}_1 \in (lb_1, upb_1)$, lo que lleva a resolver el sistema

$$\begin{cases} \mathbf{b}_1 \sum (x_i - \mathbf{q})^2 - \sum y_i (x_i - \mathbf{q}) = 0 \\ \mathbf{b}_1^2 (n\mathbf{q} - \sum x_i) + \mathbf{b}_1 \sum y_i = 0. \end{cases}$$

e) **Varianza de estimación de predicción, $f_n^{pm}(x_0; \mathbf{b}) = \mathbf{b}_{1n}^{pmo} (x_0 - \mathbf{q}_n^{pmo})$.**

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{q}_n^{pmo} \in (rtl, rt), \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) = \mathbf{s}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x}_n)^2}{\sum (x_i - \bar{x}_n)^2} \right),$$

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{q}_n^{pmo} \in (rtl, rtu), \mathbf{b}_{1n}^{pmo} = lb_1) = \frac{\mathbf{s}^2}{n},$$

$$V_n(f_n^{pm}(x_0; \mathbf{b}) | \mathbf{q}_n^{pmo} = rtl, \mathbf{b}_{1n}^{pmo} \in (lb_1, upb_1)) = \frac{\mathbf{s}^2 (x_0 - rtl)^2}{\sum (x_i - rtl)^2}.$$

De forma análoga se obtiene cuando $\mathbf{b}_{1n}^{pmo} = upb_1$ y $\mathbf{q}_n^{pmo} = rtu$.

3. PRIORI NORMAL EN q Y EN b_1 .

Consideremos que la distribución a priori sobre los parámetros $\pi_0(\theta, \beta_1)$ es tal que $q \sim N(m_q, s_q^2)$, $b_1 \sim N(m_{b_1}, s_{b_1}^2)$ y q y b_1 independientes. La densidad a priori es

$$p_0(q, b_1) \propto \exp \left\{ -\frac{1}{2} \left[\left(\frac{b_1 - m_{b_1}}{s_{b_1}} \right)^2 + \left(\frac{q - m_q}{s_q} \right)^2 \right] \right\}.$$

La distribución conjunta de Y_n , q y b_1 queda como

$$h(Y_n, q, b_1 | X_n, p_0) \propto \exp \left\{ -\frac{1}{2} \left[\frac{(Y_n - b_1 Z_n q_1)'(Y_n - b_1 Z_n q_1)}{s^2} + \frac{(b_1 - m_{b_1})^2}{s_{b_1}^2} + \frac{(q - m_q)^2}{s_q^2} \right] \right\}.$$

a) Conjunta a posteriori.

$$p_n(q, b_1) \propto \exp \left\{ -\frac{b_1^2 (s_{b_1}^2 s_q^2 q_1' Z_n' Z_n q_1 + s^2 s_q^2) - 2b_1 (s_{b_1}^2 s_q^2 Y_n' Z_n q_1 + s^2 s_q^2 m_{b_1}) + s_{b_1}^2 s^2 (q^2 - 2qm_q)}{2s^2 s_{b_1}^2 s_q^2} \right\}.$$

b) Condicionales a posteriori.

$$p_n(q | b_1) \sim N \left(\frac{s_q^2 n b_1 (b_1 \bar{x}_n - \bar{y}_n) + s^2 m_q}{s_q^2 n b_1^2 + s^2}, \frac{s^2 s_q^2}{s_q^2 n b_1^2 + s^2} \right)$$

y

$$p_n(b_1 | q) \sim N \left(\frac{s_{b_1}^2 \sum y_i (x_i - q) + s^2 m_{b_1}}{s_{b_1}^2 \sum (x_i - q)^2 + s^2}, \frac{s^2 s_{b_1}^2}{s_{b_1}^2 \sum (x_i - q)^2 + s^2} \right).$$

Éstas se obtienen desarrollando la parte del exponente de $p_n(q, b_1)$ que depende de q y b_1 :

$$\frac{b_1^2 (s_{b_1}^2 s_q^2 \sum (x_i - q)^2 + s^2 s_q^2) - 2b_1 (s_{b_1}^2 s_q^2 \sum y_i (x_i - q) + s^2 s_q^2 m_{b_1}) + s_{b_1}^2 s^2 (q^2 - 2qm_q)}{2s^2 s_{b_1}^2 s_q^2}.$$

c) Marginales a posteriori.

$$p_n(\mathbf{q}) \propto \sqrt{V_q^*} \exp \left\{ \frac{1}{2} \left(\frac{(E_q^*)^2}{V_q^*} - \frac{\mathbf{q}^2 - 2\mathbf{q}\mathbf{m}_q}{\mathbf{s}_q^2} \right) \right\}$$

$$\text{con } E_q^* = \frac{\mathbf{s}_{b_1}^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 \mathbf{m}_{b_1}}{\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2} \text{ y } V_q^* = \frac{\mathbf{s}^2 \mathbf{s}_{b_1}^2}{\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2}.$$

$$p_n(\mathbf{b}_1) \propto \sqrt{V_{b_1}^*} \exp \left\{ \frac{1}{2} \left[\frac{(E_{b_1}^*)^2}{V_{b_1}^*} - \frac{\mathbf{b}_1^2 (\mathbf{s}_{b_1}^2 \sum x_i^2 + \mathbf{s}^2) - 2\mathbf{b}_1 (\mathbf{s}_{b_1}^2 \sum x_i y_i + \mathbf{s}^2 \mathbf{m}_{b_1})}{\mathbf{s}^2 \mathbf{s}_{b_1}^2} \right] \right\}$$

$$\text{con } E_{b_1}^* = \frac{\mathbf{s}_q^2 n \mathbf{b}_1 (\mathbf{b}_1 \bar{x}_n - \bar{y}_n) + \mathbf{s}^2 \mathbf{m}_q}{\mathbf{s}_q^2 n \mathbf{b}_1^2 + \mathbf{s}^2} \text{ y } V_{b_1}^* = \frac{\mathbf{s}^2 \mathbf{s}_q^2}{\mathbf{s}_q^2 n \mathbf{b}_1^2 + \mathbf{s}^2}.$$

La derivada de la densidad marginal a posteriori de \mathbf{q} igualada a cero resulta el polinomio de quinto grado

$$\mathbf{s}_q^2 [(n\mathbf{q} - \sum x_i)(\mathbf{s}_{b_1}^2 \mathbf{s}^2 \text{aux1} + \text{aux2}^2) + \text{aux1} \text{aux2} \sum y_i] + (\mathbf{q} - \mathbf{m}_q) \text{aux1}^2 \mathbf{s}^2 = 0$$

con $\text{aux1} = \mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2$ y $\text{aux2} = \mathbf{s}_{b_1}^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 \mathbf{m}_{b_1}$, lo que indica que a lo sumo la distribución puede ser trimodal.

d) Modas a posteriori.

Las modas a posteriori se obtendrán maximizando el exponente de la densidad $p_n(\mathbf{q}, \mathbf{b}_1)$, lo que lleva a resolver el sistema

$$\begin{cases} \mathbf{b}_1 (\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2) - (\mathbf{s}_{b_1}^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 \mathbf{m}_{b_1}) = 0 \\ \mathbf{q} (n \mathbf{b}_1^2 + \mathbf{s}^2) - (\mathbf{s}_q^2 n \mathbf{b}_1 (\mathbf{b}_1 \bar{x}_n - \bar{y}_n) + \mathbf{s}^2 \mathbf{m}_q) = 0 \end{cases}$$

cuya solución analítica pasa por la resolución de polinomios de orden superior.

Una priori de características similares a la de esta sección, es la que considera que $\mathbf{q} \sim N^{[rtl,rtu]}(\mathbf{m}_q, \mathbf{s}_q^2)$ y $\mathbf{b}_1 \sim N(\mathbf{m}_{b_1}, \mathbf{s}_{b_1}^2)$. Esta pequeña variación hace que la normalidad de $\mathbf{p}_n(\mathbf{q} | \mathbf{b}_1)$ se mantenga pero truncada doblemente, y en el caso de la marginal $\mathbf{p}_n(\mathbf{q})$ aparezca la función indicatriz de \mathbf{q} sobre $[rtl,rtu]$. Para esta nueva priori, la distribución de $\mathbf{p}_n(\mathbf{b}_1 | \mathbf{q})$ permanece igual salvo la restricción de la indicatriz de \mathbf{q} y en la densidad marginal $\mathbf{p}_n(\mathbf{b}_1)$ aparece el factor $F\left(\frac{rtu - E_{b_1}^*}{\sqrt{V_{b_1}^*}}\right) - F\left(\frac{rtl - E_{b_1}^*}{\sqrt{V_{b_1}^*}}\right)$ donde

$$E_{b_1}^* = \frac{\mathbf{s}_q^2 n \mathbf{b}_1 (\mathbf{b}_1 \bar{x}_n - \bar{y}_n) + \mathbf{s}^2 \mathbf{m}_q}{\mathbf{s}_q^2 n \mathbf{b}_1^2 + \mathbf{s}^2} \text{ y } V_{b_1}^* = \frac{\mathbf{s}^2 \mathbf{s}_q^2}{\mathbf{s}_q^2 n \mathbf{b}_1^2 + \mathbf{s}^2}.$$

4. PRIORI NORMAL EN \mathbf{q} Y UNIFORME EN \mathbf{b}_1 .

Consideremos que la distribución a priori sobre los parámetros, $\pi_0(\theta, \beta_1)$, es tal que $\mathbf{q} \sim N(\mathbf{m}_q, \mathbf{s}_q^2)$, $\mathbf{b}_1 \sim U[lb_1, upb_1]$ y \mathbf{q} y \mathbf{b}_1 independientes. La densidad a priori es

$$\mathbf{p}_0(\mathbf{q}, \mathbf{b}_1) \propto I_{[lb_1, upb_1]} \exp\left\{-\frac{1}{2}\left(\frac{\mathbf{q} - \mathbf{m}_q}{\mathbf{s}_q}\right)^2\right\}.$$

La distribución conjunta de Y_n , \mathbf{q} y \mathbf{b}_1 queda como sigue

$$h(Y_n, \mathbf{q}, \mathbf{b}_1 | X_n, \mathbf{p}_0) \propto I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{-\frac{1}{2}\left[\frac{(Y_n - \mathbf{b}_1 Z_n \mathbf{q}_1)'(Y_n - \mathbf{b}_1 Z_n \mathbf{q}_1)}{\mathbf{s}^2} + \frac{(\mathbf{q} - \mathbf{m}_q)^2}{\mathbf{s}_q^2}\right]\right\}.$$

a) Conjunta a posteriori.

$$\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1) \propto I_{[lb_1, upb_1]} \exp\left\{-\frac{\mathbf{b}_1^2 \mathbf{s}_q^2 \mathbf{q}_1' Z_n' Z_n \mathbf{q}_1 - 2\mathbf{b}_1 \mathbf{s}_q^2 Y_n' Z_n \mathbf{q}_1' + \mathbf{s}^2 (\mathbf{q}^2 - 2\mathbf{q} \mathbf{m}_q)}{2\mathbf{s}^2 \mathbf{s}_q^2}\right\}.$$

b) Condicionales a posteriori.

$$\mathbf{p}_n(\mathbf{q} | \mathbf{b}_1) \sim \begin{cases} N\left(\frac{n\mathbf{b}_1 \mathbf{s}_q^2 (\mathbf{b}_1 \bar{x}_n - \bar{y}_n) + \mathbf{s}^2 \mathbf{m}_q}{n\mathbf{b}_1^2 \mathbf{s}_q^2 + \mathbf{s}^2}, \frac{\mathbf{s}^2 \mathbf{s}_q^2}{n\mathbf{b}_1^2 \mathbf{s}_q^2 + \mathbf{s}^2}\right) & \text{si } \mathbf{b}_1 \in [lb_1, upb_1] \\ 0 & \text{en el resto} \end{cases}$$

y

$$\mathbf{p}_n(\mathbf{b}_1 | \mathbf{q}) \propto I_{[lb_1, upb_1]}(\mathbf{b}_1) \exp\left\{\frac{\mathbf{b}_1^2 \sum (x_i - \mathbf{q})^2 - 2\mathbf{b}_1 \sum y_i (x_i - \mathbf{q})}{\mathbf{s}^2}\right\}.$$

Éstas se obtienen desarrollando el exponente de $\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1)$ que depende de \mathbf{q} y \mathbf{b}_1 como

$$-\frac{\mathbf{b}_1^2 (\mathbf{s}_q^2 \sum (x_i - \mathbf{q})^2 - 2\mathbf{b}_1 \mathbf{s}_q^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 (\mathbf{q}^2 - 2\mathbf{q} \mathbf{m}_q)}{2\mathbf{s}^2 \mathbf{s}_q^2}.$$

c) Marginales a posteriori.

$$\mathbf{p}_n(\mathbf{q}) \propto \sqrt{V_q^*} \left[\mathbf{F} \left(\frac{upb_1 - E_q^*}{\sqrt{V_q^*}} \right) - \mathbf{F} \left(\frac{lb_1 - E_q^*}{\sqrt{V_q^*}} \right) \right] \exp \left\{ \frac{1}{2} \left(\frac{(E_q^*)^2}{V_q^*} - \frac{\mathbf{q}^2 - 2\mathbf{q}\mathbf{m}_q}{\mathbf{s}_q^2} \right) \right\},$$

$$\text{con } E_q^* = \frac{\mathbf{s}_{b_1}^2 \sum y_i (x_i - \mathbf{q}) + \mathbf{s}^2 \mathbf{m}_{b_1}}{\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2} \text{ y } V_q^* = \frac{\mathbf{s}^2 \mathbf{s}_{b_1}^2}{\mathbf{s}_{b_1}^2 \sum (x_i - \mathbf{q})^2 + \mathbf{s}^2}.$$

$$\mathbf{p}_n(\mathbf{b}_1) \propto I_{[lb_1, upb_1]} \sqrt{V_{b_1}^*} \exp \left\{ \frac{1}{2} \left[\frac{(E_{b_1}^*)^2}{V_{b_1}^*} - \frac{\mathbf{b}_1^2 \sum x_i^2 - 2\mathbf{b}_1 \sum x_i y_i}{\mathbf{s}^2} \right] \right\},$$

$$\text{con } E_{b_1}^* = \frac{\mathbf{s}_q^2 n \mathbf{b}_1 (\mathbf{b}_1 \bar{x}_n - \bar{y}_n) + \mathbf{s}^2 \mathbf{m}_q}{\mathbf{s}_q^2 n \mathbf{b}_1^2 + \mathbf{s}^2} \text{ y } V_{b_1}^* = \frac{\mathbf{s}^2 \mathbf{s}_q^2}{\mathbf{s}_q^2 n \mathbf{b}_1^2 + \mathbf{s}^2}.$$

d) Modas a posteriori.

Se obtendrán maximizando el exponente de la densidad conjunta a posteriori $\mathbf{p}_n(\mathbf{q}, \mathbf{b}_1)$ con la restricción de que $\mathbf{b}_1 \in [lb_1, upb_1]$, lo que lleva a resolver el sistema

$$\begin{cases} \mathbf{b}_1 \mathbf{s}_q^2 \sum (x_i - \mathbf{q})^2 - \mathbf{s}_q^2 \sum y_i (x_i - \mathbf{q}) = 0 \\ \mathbf{q} (n \mathbf{b}_1^2 \mathbf{s}_q^2 + \mathbf{s}^2) - (\mathbf{s}_q^2 n \mathbf{b}_1 (\mathbf{b}_1 \bar{x}_n - \bar{y}_n) + \mathbf{s}^2 \mathbf{m}_q) = 0 \end{cases}$$

cuya solución analítica pasa por la resolución de polinomios de orden superior.

BIBLIOGRAFIA

- Abdelbasit, K.M. y Plackett, R.L. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, **78**: 90-98.
- Anderson, T.W. y Taylor, J.B. (1976). Some experimental results on the Statistical properties of least square estimates in control problems. *Econometrica*, **44**: 1289-1302.
- Astrom, K.J. y Wittenmark, B. (1995). *Adaptive Control, 2nd Edition*. Reading, Massachussets: Addison Wesley
- Atkinson, A.C. y Donev, A.N. (1992). *Optimal Experimental Designs*. Oxford: Oxford University Press.
- Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag, 2nd ed.
- Berger, J., Berliner, M. Y Zaman, A. (1982). General Admissibility and inadmissibility results for estimation in a control problem. *Annals of Statistics*, **7**: 838-856.
- Berliner, M. (1982). Uniform improvements on the certainty equivalent rule in a statistical control problem. *Statistical Decision Theory and Related Topics III, Vol. 1*,(eds. S.S. Gupta and J. Berger), pp. 157-168. New York: Academic Press.
- Berliner, M. (1983). Improving on inadmissible estimators in control problem. *Annals of Statistics*, **11**: 814-826.
- Berliner, M. (1987). Bayesian control in mixture models. *Technometric*, **29**: 455-460.
- Bernardo, J.M. y Smith, A.F.M. (1994). *Bayesian Theory*. New York: Wiley
- Berry, D.A. y Eick, S.G. (1995). Adaptive assignment versus balanced randomization in clinical trials, a decision-analysis. *Statistical in Medicine*, **14**: 231-246.
- Berry, D.A. y Fristedt, B. (1985). *Bandit problems: sequential allocation of experiments*. Chapman and Hall, London.
- Berry, D.A. y Stangl, D.K. (1996). *Bayesian Biostatistics*, vol. 151 of *Statistics: Textbooks and Monographs*. New York: Marcel Dekker.
- Berry, D. (1993). A case for Bayesianism in clinical trials (with discussion). *Statistics in Medicine*, **12**: 1377-1404.
- Boscardin, W.J. y Gelman, A. (1996). Bayesian regression with parametric models for heteroscedasticity. *Advances in Econometrics* **11A**, 87-109.

- Box, G.E.P. (1988). Signal-to-noise ratios, performance criteria and transformations (with discussion). *Technometrics*, **30**: 1-40.
- Box, G.E.P. y Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison Wesley.
- Brent, R.P. (1973). *Algorithms of Minimization without Derivatives*. Englewood Cliffs, NJ: Prentice Hall).
- Brown, L.D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Annals Mathematics Statistical*, **37**: 1087-1136.
- Buonaccorsi, J.P. y Iyer, H.K. (1986). Optimal design for ratios of linear combinations in the general linear model. *Journal of Statistical Planning and Inference*, **13**: 345-356.
- Buonaccorsi, J.P. (1986). Design considerations for calibration. *Technometrics*, **28**: 149-155.
- Chaloner, K. (1989). Bayesian design for estimating the turning point of a quadratic regression. *Communications in Statistics; Theory and Methods*, **18**: 1385-1400.
- Chaloner, K. (1991). Bayesian residual analysis in the presence of censoring. *Biometrika*, **78**: 637-644.
- Chaloner, K. y Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika*, **75**: 651-659.
- Chaloner, K. y Verdinelli, I. (1996). Bayesian experimental design: A review. *Statistical Science*, **11**: 273-304.
- Chen, K. y Hu, I. (1998). On consistency of Bayes estimates in a certainty equivalence adaptive system. *IEEE Transactions on Automatic Control*, **43**: 943-947.
- De Groot, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58**: 403-417.
- Fedorov, V.V. (1972). *Theory of Optimal Experiments*. New York: Academic Press.
- Fieller, E.C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society, (Series B)*, **16**: 175-185.
- Fieller, F.C. (1940). The biological standardization of insulin. *Journal Royal Statistical Society Supply*, **7**: 1-64.
- Ford, I., Kitsos, C.P. y Titterton, D.M. (1989). Recent advances in non-linear experimental designs. *Technometrics* **31**: 49-60.
- Frees, E.W. y Ruppert, D. (1990). Estimation following a Robbins-Monro designed experiment. *Journal of the American Statistical Association*, **85**: 1123-1129.
- Gelman, A., Carlin, J.B., Stern, H.S. y Rubin, D.B. (1995). *Bayesian Data Analysis*. Chapman y Hall.

-
- Gill, P.E., Murray, W. y Wright, M.H. (1981). *Practical Optimization*. New York: Academic Press.
 - Ginebra, J. (1996). Small sequential designs that stay close to a target. *COMPSTAT 1996. Proceedings in Computational Statistics*, (ed. A. Prat), pp. 217-276. Heidelberg: Physica Verlag.
 - Ginebra, J. y Clayton, M.K. (1995). The response surface bandit. *Journal of the Royal Statistical Society, (Series B)*, **57**: 771-784.
 - Golub, G.H. y Van Loan, C.F. (1983). *Matrix Computations*. Baltimore, Maryland: Johns Hopkins University Press.
 - Graybill, F.A. (1976). *Theory and application of the general model*. Massachusetts: Duxbury Press.
 - Hardwick, J.P. y Stout, Q.F. (1998). Flexible algorithms for creating and analyzing adaptive sampling procedures. *New Developments and Applications in Experimental Design*.
 - Hu, I. (1997). Strong consistency in stochastic regression models via posterior covariance matrices. *Biometrika*, **84**: 744-749.
 - Hu, I. (1998). On sequential designs in nonlinear problems. *Biometrika*, **85**: 496-503.
 - Kiefer, J.C. y Wolfowitz, J. (1959). Optimum designs in regression problems. *Annals of Mathematical Statistics*, **30**: 271-294.
 - Knuth, D.E. (1968). *Fundamental Algorithms, vol. 1 of The Art of Computer Programming*. MA: Addison-Wesley.
 - Lai, T.L. (1987). Adaptive treatment allocation and the multiarmed bandit problem. *Annals of Statistics*, **15**: 1091-1114.
 - Lai, T.L. y Robbins, H. (1979). Adaptive design and stochastic approximation. *The Annals of Statistics*, **7**: 1196-1221.
 - Lai, T.L. y Robbins, H. (1982). Adaptive design and the multiperiod control problem. *Statistical Decision Theory and Related Topics III, Vol. 2*, (eds. S.S. Gupta and J. Berger), pp. 103-120. New York: Academic Press.
 - Lai, T.L. y Wei, C.Z. (1982). Least square estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistical*, **10**: 154-166.
 - McLeish, D.L. y Tosh, D. (1983). The estimation of extreme quantiles in logit bioassay. *Biometrika*, **70**: 625-632.
 - McLeish, D.L. y Tosh, D. (1990). Sequential designs in bioassay. *Biometrics*, **46**: 103-116.
 - Müller, P. y Parmigianni, G. (1993). Numerical evaluation of information theoretic measures. Duke University, ISDS, Technical Report 93-A05.

- Müller, P., Berry, D., Grieve, A., Smith, M. Y Krams, M. (1999). Simulation-based sequential Bayesian design. Technical Report, University Duke.
- Nelder, J.A. y Mead, R. (1965). *Computer Journal*, **7**: 308-313.
- Neter, J., Wasserman, W. y Kutner, M. (1989). *Applied Linear Regression Models*, 2nd edition. Homewood, Illinois: Richard D. Irwin, Inc.
- Prescott, E. C. (1972). The multiperiod control problem under uncertainty. *Econometrica*, **40**: 1043-1058.
- Prescott, E.C. (1974). Adaptive decision rules for macroeconomic planning. *Studies in Bayesian Econometrics and Statistics*, (eds. S.E. Fienberg and A. Zellner). North Holland.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. y Flannery, B.P. (1992). *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press.
- Pulkelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.
- Putterman, M.L. (1994). *Markov Decision Processes*. Wiley, New York.
- Robbins, H. y Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, **22**: 400-407.
- Robbins, H. y Siegmund, D. (1974). Sequential tests involving two populations. *J. Am. Statist. Assoc.* **69**: 132-139.
- Robert, C.P. (1994). *The Bayesian Choice: A Decision-Theoretic Motivation*. Springer Texts in Statistics.
- Rootzen, H. y Sternby, J. (1984). Consistency in least squares: A Bayesian approach. *Automatica* **20**: 471-475.
- Rosenberger, W. (1996). New directions in adaptive designs. *Statistical Science*, **11**: 137-149.
- Ruppert, D. (1991). *Stochastic approximation*. In *Handbook of Sequential Analysis* (eds. B.K. Ghosh an P.K. Sen), pp. 503-529. Marcel Dekker, New York.
- Silvey, S.D. (1980). *Optimal Design*. London: Chapman Hall.
- Sitter, R.R. y Wu, C.F.J. (1993). Optimal designs for binary response experiments: Fieller, D and A criteria. *Scandinavian Journal of Statistics*, **20**: 329-341.
- Spiegelhalter, D.J., Freedman, L.S. y Parmar, M.K.B. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society, Series A*, **157**: 357-387.
- Srinivasan, C. (1984). A sharp necessary and sufficient condition for inadmissibility of estimators in a control problem. *Annals of Statistics*, **12**: 927-944.
- Tanner, M.A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Series in Statistics.
- Tewarson, R.P. (1973). *Sparse Matrices*. New York: Academic Press.

-
- Verdinelli, I. y Wynn, H.P. (1988). Targer attainment and experimental design, a Bayesian approach. In *Optimal Design and Analysis of Experiments*, (eds. Y. Dodge, V.V. Fedorov and H.P. Wynn), North-Holland, Amsterdam.
 - Weisberg, S. (1985). *Applied Linear Regression*, 2nd edition. New York: Wiley.
 - Wu, C.J.F. (1986). Maximum likelihood recursion and stochastic approximation in sequential designs. *Adaptive Statistical Procedures and Related Topics* (ed. J. Van Ryzin), pp. 298-313. Institute of Mathematical Statistics.
 - Wu, C.J.F. (1988). Optimal design for percentile estimation of a quantal response curve. In *Optimal Design and Analysis of Experiments*, (eds, Y. Dodge, V.V. Fedorov and H.P. Wynn), pp. 213-223. Amsterdam: North Holland.
 - Ying, Z. Y Wu, C.F.J. (1997). An Asymptotic theory of sequential designs based on maximum likelihood recursions. *Statistica Sinica*, **7**: 75-91.
 - Zacks, S. Y Eichhorn, B.H. (1975). Sequential Search of Optimal Dosages: The Linear Regression Case. *A Survey of Statistical Design Design and Linear Models*. North-Holland.
 - Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. J. Wiley, New York.
 - Zellner, A. (1975). Bayesian analysis of regression error terms. *Journal of the American Statistical Association*, **70**: 138-144.