

Capítulo 6

Desarrollos: primera parte

Una vez que ya hemos presentado el plan de estudio que nos proponemos abordar, comenzaremos a evaluar la Etapa I que definimos en el capítulo anterior. Para ello, partimos de un caso de estudio, que toma como inicio a una terna de valores fijos de las 3 variables de estudio — w, L y S (factor de centro, lado del diseño y salto entre centros, respectivamente)— para el que desarrollaremos el ajuste secuencial de modelos de acuerdo con la estrategia que hemos elegido seguir: partir de un presupuesto fijo de experimentos y ajustar los modelos de acuerdo con el criterio jerárquico de selección de términos. Al llegar al modelo final para nuestro caso —en el que se haya utilizado todo el presupuesto disponible— evaluaremos en el capítulo siguiente la extensión a la Etapa II, en la que consideraremos los 15 casos de puntos experimentales.

6.1. Introducción

Siguiendo el mapa de estudio que nos propusimos en el capítulo anterior, en este capítulo abordaremos la *Etapa I*, en la que dejaremos fijos los valores de las 3 variables de estudio — L, S , y w — y con las que generaremos un “caso” de puntos de diseño partiendo de una misma superficie teórica¹. Veremos que inicialmente este caso comprenderá 729 modelos posibles. Luego de ciertas depuraciones que se le aplicarán para seleccionar solamente ciertos casos particulares, el mismo quedará reducido a 60 modelos posibles. Los detalles de todos los programas que hemos realizado se encuentran descritos en el **Apéndice C**.

Puesto que partimos de una superficie teórica de la que no disponemos puntos de funcionamiento del proceso real, proponemos aquí una forma de definir un punto de funcionamiento de valor conocido², que utiliza la misma superficie simulada para

¹En este contexto, diremos que el valor β que se escoge para definir la superficie teórica sería el primer “input” que es necesario definir.

²La diferencia que en el fondo existe entre la generación del primer punto de diseño y la consi-

definirlo, y que también nos servirá para definir una lista de valores posibles de las 3 variables de estudio, los cuales iremos evaluando en los dos primeros capítulos de la **Parte III (Capítulos 8 y 9)**.

La propuesta que presentamos aquí es de carácter general, aplicable a procesos que utilicen respuestas del tipo binario con 2 factores de control. En este capítulo lo ilustraremos suponiendo un proceso real mediante simulación.

6.2. Estudio de un caso

6.2.1. Introducción

Como hemos comentado, la superficie de respuesta que simularemos representará nuestro proceso generador de datos, la que llamaremos de ahora en más *superficie teórica*, abreviada como *ST*. Comentaremos en ésta y en las secciones restantes del capítulo la metodología que hemos adoptado para generar las *ST* que nos proporcionarán las observaciones de la respuesta, con las que ajustaremos modelos de manera secuencial. Ya que este proceso lo hemos implementado en forma de un programa *ad-hoc* que corre bajo **R**, nos ceñiremos a la descripción de los puntos más importantes del mismo, comentándolos en los casos que corresponda. En este caso, el programa correspondiente es de la serie **BETAS** (ver **Apéndice C**).

6.2.2. Generación de datos

En los programas de la serie **BETAS**, todos los 729 modelos quedan guardados en la matriz **M1**, en cuyas columnas se pueden encontrar los valores respectivos que van tomando cada uno de los parámetros del modelo completo. Así, diremos que la matriz **M1** será la que contiene los 729 modelos posibles a partir de considerar 3 niveles para cada uno de los 6 parámetros tenidos en cuenta al generar la superficie de respuesta teórica por medio de un modelo completo de segundo orden. En nuestro caso, elegimos representar los niveles -2, 0 y 2 para dicho vector, aunque el mismo puede variarse sin problemas en el código del programa para evaluar cualquier caso que se desee.

Casos que corresponden a máximos

Puesto que cada fila de la matriz **M1** se trata de un vector de parámetros con el que se obtienen funciones logísticas de dos variables, los modelos formados a partir de cada una de dichas filas pueden representar tanto superficies que contengan máximos,

deración de un proceso real es que en este último cualquier punto de funcionamiento puede servir como primer centro, ya que no se requieren más consideraciones que la de conocer el valor de la respuesta en ese punto.

mínimos y puntos de silla como así también planos. Nuestro objetivo es el de explorar superficies que contienen máximos³, con lo cual un primer paso natural es el de seleccionar aquellas configuraciones de M1 que representen superficies con máximos.

Para ello, utilizaremos el criterio de clasificar las superficies mediante los valores propios o autovalores de las matrices de términos de segundo orden, para cada uno de los 729 casos. Es posible ver más adelante en el mismo **Apéndice C**, que los casos que corresponden a máximos se reducen en total de 729 a 81 filas de la misma matriz M1, para lo cual hemos utilizado las líneas de comando correspondientes a los MÓDULOS 2 y 3 del mismo programa BETAS.

La matriz que contiene los modelos que corresponden solamente a máximos queda identificada como M4, que tiene la misma estructura que la M1 con el agregado de dos columnas Lam.1 y Lam.2, que corresponden a los valores propios, todos con signos negativos, que son indicativos de superficies que contienen de máximos.

Caracterización de los puntos máximos

Mediante la utilización de los MÓDULOS 4 y 5 del mismo programa, se pueden calcular las coordenadas en el plano de los factores de los puntos máximos de cada uno de los 81 modelos correspondientes a las filas de la matriz M4, como así también el valor que toma la función $\pi(\mathbf{x}, \beta)$ en dichos puntos. El resultado final es la matriz M6, en donde quedan evidenciados los valores precitados para cada uno de los modelos.

Depuración de la matriz de modelos

Al disponer de un recurso tan útil y flexible como lo es el lenguaje **R**, podemos “depurar” la matriz de modelos M6 de tal forma que, previendo futuras exploraciones de las superficies de respuesta correspondientes, se eliminen todos aquellos casos en donde el máximo de la superficie coincida con el origen de coordenadas, es decir, con el punto $\mathbf{x} = (0, 0)'$. El motivo por el cual hemos considerado esta depuración radica en razones de buen sentido: al querer buscar las condiciones operativas de variables de control concretas que nos lleven a un máximo en el valor esperado de la respuesta, no tiene demasiado sentido indicar que el valor esperado del máximo se conseguirá en el punto $\mathbf{x} = (0, 0)'$. En la práctica, esto equivale a indicar que el valor esperado de la respuesta alcanzará su máximo cuando los factores estén evaluados en su valor nulo, lo cual no parece tener demasiado sentido. A tal efecto, mediante la ejecución del MÓDULO 7, llegamos a la matriz M7, que es la M6 depurada, cuyo número de filas se reducirá de 81 a 60.

³Y esto es debido al tipo de problema que estudiamos: maximizar una probabilidad de éxito mediante el control de los factores de variabilidad que presumiblemente influyan sobre ésta.

Por lo tanto, la matriz `M7` será la que contenga los 60 modelos que tendremos disponibles para el estudio de superficies de respuesta que contengan proporciones. Esta matriz será nuestra “base de datos” en donde se almacenan todos los casos que contienen máximos. Cada una de sus filas será considerada como un “proceso generador” de observaciones, que representará el modelo o proceso de la realidad que nos proporciona las observaciones con las cuales estudiaremos nuestro problema: la exploración experimental del mismo mediante una estrategia secuencial de diseños. Una de estas filas será la que elegiremos para representar nuestro proceso de la realidad, con el que evaluaremos distintas alternativas para explorar su estructura y ver si podemos acercarnos a mejores condiciones operativas.

Un resumen de la matriz `M7` puede verse a continuación, cuyo número de filas está encabezado por la columna `ID`:

```
> M7
      ID B.00 B.01 B.02 B.11 B.22 B.12 Lam.1 Lam.2  X1.S  X2.S  Pi.Max
1      1  -2  -2  -2  -2  -2  -2  -1  -3 -0.3333 -0.3333 0.20861
2      2   0  -2  -2  -2  -2  -2  -1  -3 -0.3333 -0.3333 0.66076
3      3   2  -2  -2  -2  -2  -2  -1  -3 -0.3333 -0.3333 0.93503
...
511 58  -2   2   2  -2  -2   2  -1  -3  1.0000  1.0000 0.50000
512 59   0   2   2  -2  -2   2  -1  -3  1.0000  1.0000 0.88080
513 60   2   2   2  -2  -2   2  -1  -3  1.0000  1.0000 0.98201

> dim(M7)
[1] 60 12
```

6.2.3. Selección de un caso

Nuestro punto de partida para este capítulo será el de elegir un modelo apropiado⁴ a partir de una de las filas de la matriz `M7`. Definimos en **R** una variable llamada `k.2`, que escoge al azar una fila —es decir, un modelo particular— de la matriz `M7`. Al vector fila resultante, lo guardamos en la variable `b.k.2`.

```
> k.2 <- sample(M7[, "ID"], 1)
```

⁴Queda claro que el experimentador tendrá que contar con conocimiento complementario a su proceso —y no sólo intrínseco— para determinar las características de esta superficie teórica. Nosotros hemos representado este proceso de elección del modelo adecuado mediante una selección al azar de una fila de la matriz `M7`, que contiene los 60 modelos disponibles, aunque el experimentador elegiría el modelo más adecuado.

```
> b.k.2 <- M7[k.2,]
```

```
b.k.2
```

```
   ID B.00 B.01 B.02 B.11 B.22 B.12 Lam.1 Lam.2  X1.S  X2.S Pi.Max
495 45    2    2   -2   -2   -2    2    -1    -3 0.3333 -0.3333 0.93503
```

Este resultado corresponde entonces a la fila identificada con el número 495 (índice de la matriz M1) o bien con el 45 (índice de la matriz M7). Así, leyendo adecuadamente este vector fila, el modelo teórico simulado en escala logística tiene las siguientes características:

a. Valor del vector de parámetros del modelo:

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_{12}, \beta_{11}, \beta_{22})' = (2, 2, -2, -2, -2, 2)'$$

b. Superficie teórica completa de segundo orden en escala logística:

$$\pi(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp(2 + x_1 - 2x_2 - 2x_1^2 - 2x_2^2 + 2x_1x_2)}{1 + \exp(2 + x_1 - 2x_2 - 2x_1^2 - 2x_2^2 + 2x_1x_2)} \quad (6.1)$$

c. Forma matricial del predictor lineal del modelo:

$$\pi(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp(\eta)}{1 + \exp(\eta)}, \quad \text{en donde: } \eta = \beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x},$$

$$\beta_0 = 2, \quad \mathbf{b} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -2 & \frac{1}{2} \cdot (-2) \\ \frac{1}{2} \cdot (-2) & -2 \end{bmatrix} = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}$$

d. Caracterización de la superficie mediante valores propios⁵:

$$\lambda_1 = -1,0, \quad \lambda_2 = -3,0 \quad (\text{autovalores correspondientes a } \boldsymbol{\beta})$$

e. Coordenadas del punto estacionario:

$$\mathbf{x}_S = \begin{bmatrix} x_{1S} \\ x_{2S} \end{bmatrix} = \begin{bmatrix} 0,3333 \\ -0,3333 \end{bmatrix} = \mathbf{x}_{\text{máx}} \quad (\text{punto estacionario correspondiente al máximo})$$

⁵ Vid. Apéndice A.

f. Valor máximo de la función:

$$\pi(\mathbf{x}_{\text{máx}}, \boldsymbol{\beta}) = 0,93503$$

g. Gráficas de la superficie y sus curvas de nivel:

En los programas de la serie MEDCA, en los MÓDULOS 1 y 2, definimos las líneas de comando, con parámetros convenientemente elegidos, que nos permiten visualizar el aspecto que tendrán tanto la superficie teórica como sus curvas de nivel para la fila que hemos seleccionado de la matriz M7. Para nuestro caso de estudio, el mismo queda resumido mediante la variable `b.k.2` del mismo programa. En el **Apéndice C** podrán verse al detalle estas instrucciones. Las gráficas obtenidas pueden verse en las figuras 6.1 y 6.2. De este modo, ya podemos tener una primera apreciación de la superficie que representa el proceso generador de observaciones de la respuesta, que corresponde a una superficie teórica completa de segundo orden elegida al azar de un conjunto de 60 configuraciones posibles y adecuadas para nuestro estudio.

h. Cálculo de $\pi(\mathbf{x}, \boldsymbol{\beta})$ para un valor dado de \mathbf{x} :

En el MÓDULO 8 de los programas de la serie BETAS hemos agregado unas líneas de programa que permiten seleccionar una fila de la matriz M7 —esto es, seleccionar un modelo particular de los 60 disponibles— y calcular con ellas el valor $\pi(\mathbf{x}, \boldsymbol{\beta})$ que le corresponde para un par de valores $\mathbf{x} = (x_1, x_2)'$. Esta función será identificada mediante el nombre `f.02`. Primero se elige una fila de la matriz y se la guarda en la variable `k.2`. Luego de ello, se elige un punto del plano de los factores y se lo evalúa haciendo `f.02(x1, x2)`. Por ejemplo, si tomamos el mismo modelo que anteriormente, comprobamos que las coordenadas del punto máximo evaluadas en `f.02` coinciden con la presentada en la variable `b.k.2`:

```
> k.2 <- 45

> M7[45, c("X1.S", "X2.S", "Pi.Max")]
      X1.S    X2.S  Pi.Max
495 0.3333 -0.3333 0.93503
```

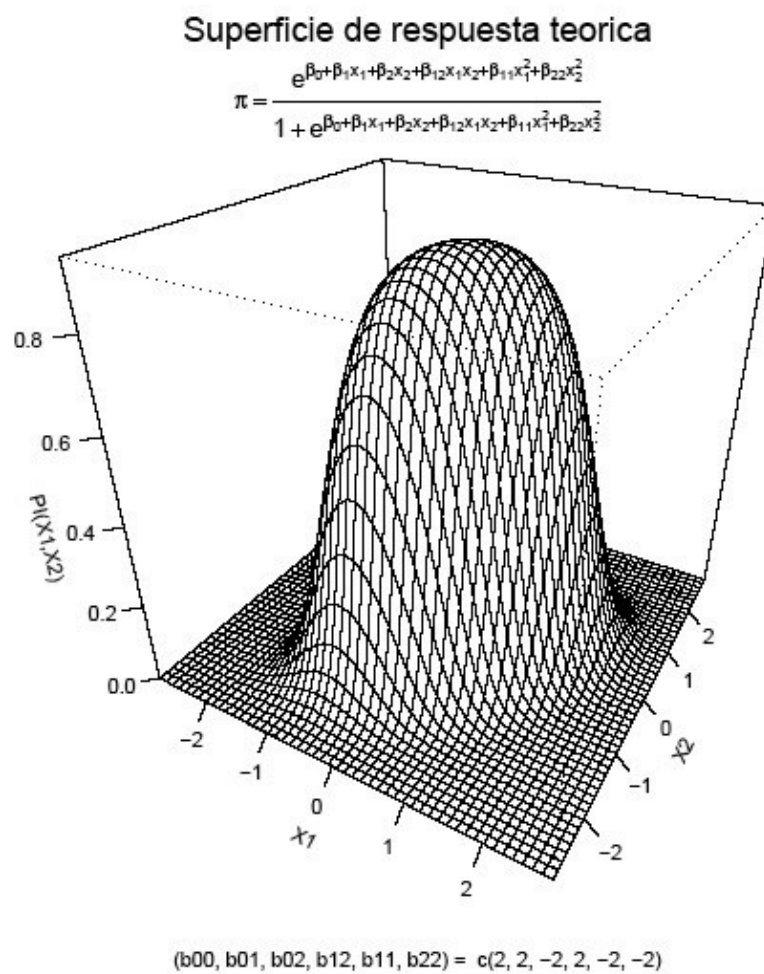


Figura 6.1: Superficie teórica en escala logística correspondiente al modelo teórico

$$\pi(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp(2+x_1-2x_2-2x_1^2-2x_2^2+2x_1x_2)}{1+\exp(2+x_1-2x_2-2x_1^2-2x_2^2+2x_1x_2)}.$$

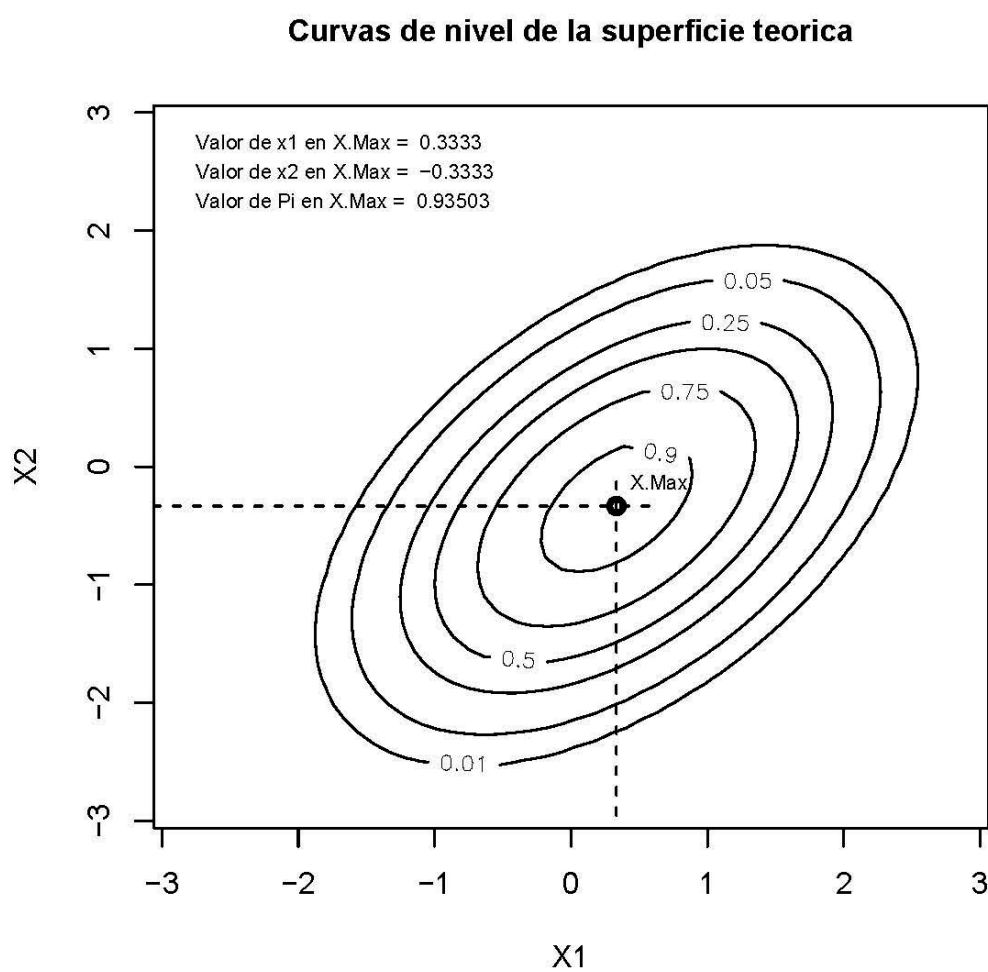


Figura 6.2: Curvas de nivel de la superficie teórica en escala logística correspondiente al modelo teórico $\pi(\mathbf{x}, \beta) = \frac{\exp(2+x_1-2x_2-2x_1^2-2x_2^2+2x_1x_2)}{1+\exp(2+x_1-2x_2-2x_1^2-2x_2^2+2x_1x_2)}$.


```
> f.02(0.3333, -0.3333)
[1] 0.935031
```

```
> f.02(M7[45,"X1.S"],M7[45,"X2.S"])
[1] 0.935031
```

Partimos del estudio de la superficie teórica generada a partir de la expresión que detallamos anteriormente en la ecuación (6.1):

$$\pi(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp(2 + x_1 - 2x_2 + 2x_1x_2 - 2x_1^2 - 2x_2^2)}{1 + \exp(2 + x_1 - 2x_2 + 2x_1x_2 - 2x_1^2 - 2x_2^2)}$$

Esta superficie será nuestro proceso generador de datos para todo par ordenado $\mathbf{x} = (x_1, x_2)'$ del plano de los factores. En particular, también tendremos en consideración el estudio del predictor lineal del modelo, dado por el exponente de la función exponencial:

$$\eta = 2 + x_1 - 2x_2 + 2x_1x_2 - 2x_1^2 - 2x_2^2 \quad (6.2)$$

A partir de esta definición, nos proponemos disponer de un presupuesto equivalente a 1500 puntos provenientes de experimentos de Bernoulli, en los que observaremos el número de éxitos y las proporciones de éstos, con los que ajustaremos modelos secuencialmente. Luego de haber utilizado todo el presupuesto disponible, evaluaremos la calidad de los ajustes mediante los criterios que indicamos en el capítulo anterior.

6.2.4. Localización del primer centro de experimentación

Una primera cuestión natural que surge al iniciar el estudio de la *MSR* versa acerca de dónde localizar los centros de diseño dentro del plano de los factores. Dado que nuestro proceso generador de datos o superficie teórica ha sido simulado y no es posible definir un “punto natural de funcionamiento” del proceso, con lo cual no podemos establecer un punto “real” en donde sepamos que el proceso puede funcionar. Por esto, nos vemos obligados a proponer una estrategia *ad-hoc* para salvar este obstáculo. Proponemos entonces, como comienzo del “aprendizaje de los datos”, un punto de funcionamiento que se encuentre sobre una de las curvas de nivel de la superficie de respuesta.

Tomamos como punto de partida el valor correspondiente a una fracción del valor esperado de la respuesta en el máximo de la superficie supuesta, el cual quedará representado por la variable que dimos en llamar “factor de centro”, w . Esta variable define la fracción que denotaremos como $\pi_{w\% \text{máx}}$, $0 < w < 100\%$, es decir:

$$\pi_{w\% \text{máx}} = w \cdot \pi_{\text{máx}}$$

Como indicamos en el capítulo anterior, tomaremos como niveles de esta variable a los 5 primeros múltiplos de 0.05^6 . Para el desarrollo de este caso que nos ocupa, consideraremos la curva de nivel que corresponde al plano que corta a la superficie a la altura del 5 % de $\pi_{\text{máx}}$, valor constante que denotaremos como $\pi_{5\% \text{ máx}} = \text{cte.}^7$. En las figuras siguientes (6.3) y (6.4) podemos ver una ilustración de lo precitado:

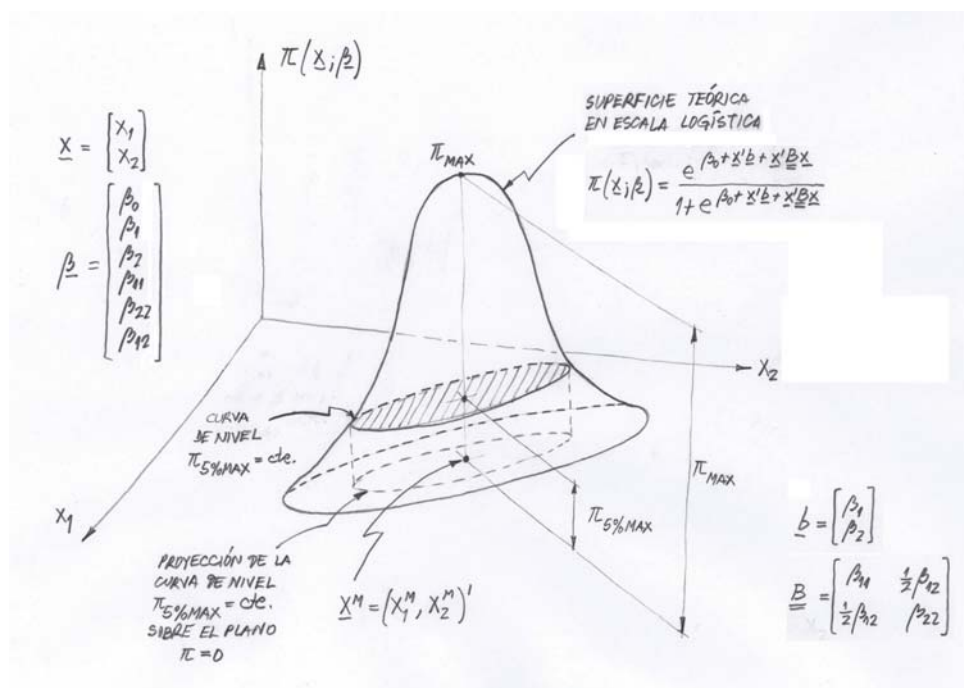


Figura 6.3: Esquema de la elección de una curva de nivel de referencia sobre la cual definir el primer centro de experimentación.

Para definir un punto relativamente sistemático que sirva para varios casos de estudio, y valiéndonos del conocimiento del punto máximo de la superficie simulada, definiremos un par de “puntos-candidatos” a ocupar el lugar del primer centro de experimentación. Dichos puntos son los que quedan definidos al cortarse la elipse con la recta mencionada, de los cuales elegiremos uno cualquiera de ellos debido a las condiciones naturales de simetría.

Tomaremos como primer par de puntos-candidato a aquellos resultantes de la in-

⁶Los motivos de la elección de estos valores responden a que mediante ellos es posible situarnos en las mismas condiciones de las que se parte en la *MSR* clásica: suponer que al comienzo de la experimentación nos encontramos en una zona “relativamente lejana” del óptimo. Considerando valores de π del 5 %, 10 %, 15 %, 20 % y 25 %, es posible encontrarnos en esta situación dentro de la *ST*: utilizando estos valores, nos aseguramos que las curvas de nivel de la *ST* resulten relativamente lejanas del óptimo.

⁷En el capítulo 8 evaluaremos al mismo tiempo un conjunto de 5 niveles para esta variable.

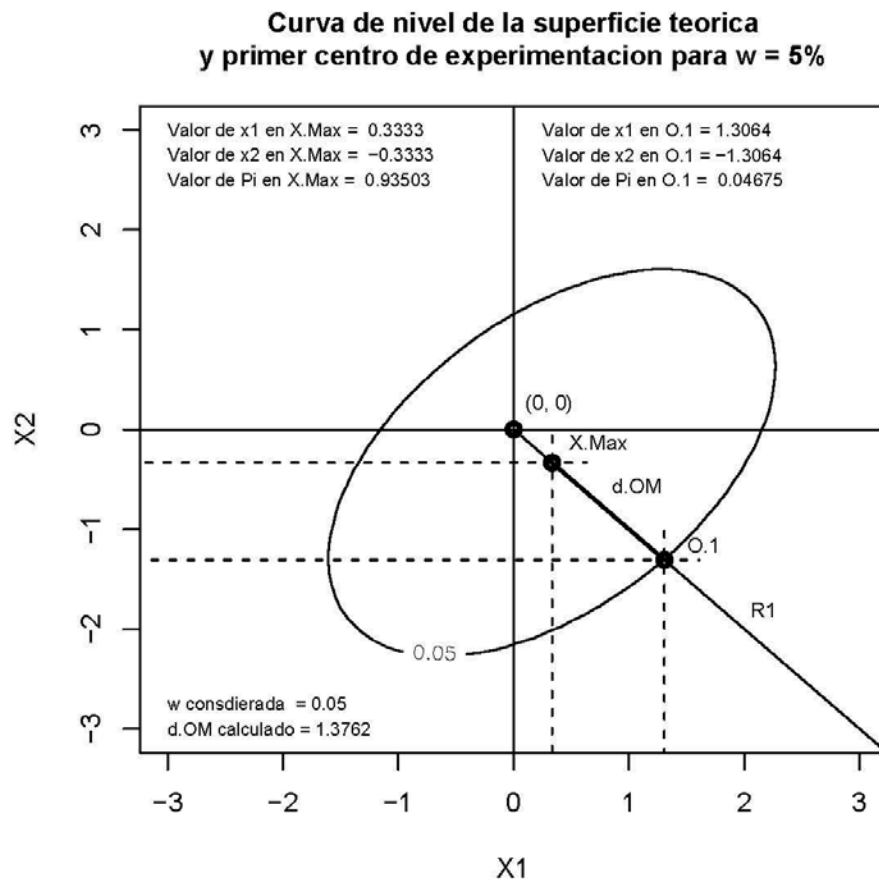


Figura 6.4: Localización de la curva de nivel $\pi_{5\% \text{ máx}} = \text{cte.}$ sobre el plano de los factores.

tersección de la curva de nivel $\pi_{5\% \text{ máx}}$ con la recta que une, en el plano de los factores, al punto máximo $\mathbf{x}^M = (x_1^M, x_2^M)'$ con el origen de coordenadas, es decir, con el punto $(x_1, x_2)' = (0, 0)'$. Denotamos mediante R_1 a esta recta y mediante \mathbf{A}^I y \mathbf{B}^I a los dos puntos en donde R_1 corta la curva de nivel, que en realidad resultarán simétricos con respecto al punto máximo de la superficie teórica.

Determinación analítica de los puntos \mathbf{A}^I y \mathbf{B}^I

Partimos del mismo modelo que definimos anteriormente, es decir:

ID	B.00	B.01	B.02	B.11	B.22	B.12	Lam.1	Lam.2	X1.S	X2.S	Pi.Max	
495	45	2	2	-2	-2	-2	2	-1	-3	0.3333	-0.3333	0.93503

Para encontrar las coordenadas de ambos puntos \mathbf{A}^I y \mathbf{B}^I , definiremos la curva de nivel de la que parte nuestro estudio. Considerando, por ejemplo, $w = 5\%$, tendremos que la curva de nivel correspondiente tendrá una ordenada igual a $w \cdot \pi_{\text{máx}}$, de valor:

$$0,05 \cdot \pi_{\text{máx}} = 0,05 \cdot 0,93503 = 0,0467515$$

$$\pi_{5\% \text{ máx}} = 0,0467515$$

Este último valor es el que nos definirá el plano paralelo al de los factores que cortará a la superficie teórica y que definirá la curva de nivel sobre la cual determinaremos el primer centro de experimentación.

La ecuación de la curva de nivel correspondiente al valor constante 0,0467515, quedará definida en función de los dos factores x_1 y x_2 mediante la expresión⁸:

$$\frac{\exp(\beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x})}{1 + \exp(\beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x})} = 0,0467515, \quad (6.3)$$

que resulta equivalente reescribirla como:

$$\beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x} = \text{logit}(0,0467515) = -3,015029$$

En términos generales, para cualquier modelo que definamos a partir de la matriz de modelos, podrá escribirse:

$$\beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x} = \ln \left(\frac{\pi_{5\% \text{ máx}}}{1 - \pi_{5\% \text{ máx}}} \right) = \text{logit}(\pi_{5\% \text{ máx}}),$$

es decir:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 = -3,015029 \quad (6.4)$$

⁸Por motivos de claridad, hemos omitido el subíndice del vector de coeficientes correspondiente a la fila elegida de la matriz de modelos.

Por otra parte, la pendiente de la recta R_1 queda definida a partir del punto máximo y del origen de coordenadas, ambos conocidos en nuestra superficie teórica, que resulta en términos generales:

$$m_{R1} = \frac{x_2^M - 0}{x_1^M - 0} = \frac{x_2^M}{x_1^M}$$

$$x_2^M = m_{R1} \cdot x_1^M \quad (6.5)$$

en la que el valor m_{R1} lleva el signo que le corresponde en función de la ubicación del punto máximo en el plano de los factores, es decir, de $\mathbf{x}^M = (x_1^M, x_2^M)'$.

Agrupando las ideas, nuestro problema impone encontrar una solución para el siguiente sistema de ecuaciones, dado por las (6.4) y (6.5):

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 = \text{logit}(\pi_5 \% \text{máx})$$

$$x_2 = m_{R1} \cdot x_1,$$

en las cuales todos los coeficientes β_j resultan conocidos, como así también las cantidades $\text{logit}(\pi_w \% \text{máx})$ y m_{R1} .

Reemplazando el valor de la variable x_2 (dado por la segunda ecuación) dentro de la primera, y agrupando términos, se llega a la siguiente ecuación cuadrática en x_1 :

$$\beta_0 - \text{logit}(\pi_5 \% \text{máx}) + (\beta_1 + m_{R1}\beta_2)x_1 + [\beta_{11} + (m_{R1})^2\beta_{22} + m_{R1}\beta_{12}]x_1^2 = 0. \quad (6.6)$$

Llamando:

$$\begin{aligned} \text{cc.01} &= \beta_0 - \text{logit}(\pi_5 \% \text{máx}) \\ \text{bb.01} &= (\beta_1 + m_{R1}\beta_2) \\ \text{aa.01} &= \beta_{11} + (m_{R1})^2\beta_{22} + m_{R1}\beta_{12}, \end{aligned}$$

mediante la función `polyroot()` de R podemos calcular las raíces de (6.6) escribiendo:

```
> polyroot(c(cc.01, bb.01, aa.01))
```

El resultado de evaluar esta función de esta manera nos conduce a las raíces `r1.x1.01` y `r2.x1.01`, soluciones en x_1 de la ecuación (6.6). Reemplazando luego éstos en la (6.5), obtenemos los valores solución en x_2 , que denotaremos como `r1.x2.01` y `r2.x2.01`.

De este modo, deducimos que las coordenadas de los puntos \mathbf{A}^I y \mathbf{B}^I , resultantes de la intersección de la recta R1 con la curva de nivel correspondiente a $\pi_{5\% \text{máx}} = \text{cte.}$, quedarán definidas (en notación del programa **R**) por:

$$\begin{aligned}\mathbf{A}^I &= \mathbf{A}.01 = (\mathbf{r1.x1.01}, \mathbf{r1.x2.01})' \\ \mathbf{B}^I &= \mathbf{B}.01 = (\mathbf{r2.x1.01}, \mathbf{r2.x2.01})'\end{aligned}$$

Ejecutando las líneas de programa correspondientes al MÓDULO 8 de la serie de programas BETAS (ver **Apéndice C**), para el modelo que hemos elegido anteriormente, y considerando el mismo valor $w = 5\%$ para la curva de nivel de referencia, llegamos a que ambas raíces de x_1 son reales, que conduce a determinar las coordenadas de ambos puntos de forma automática:

$$\begin{aligned}\mathbf{A}^I &= (-0,639780, 0,639780)' \\ \mathbf{B}^I &= (1,306447, -1,306447)'\end{aligned}$$

En el mismo MÓDULO 8 del programa realizamos una pequeña prueba, la cual nos permite concluir que ambos puntos pertenecen efectivamente a la curva de nivel correspondiente a $\pi_{5\% \text{máx}} = \text{cte.}$

Elección del primer centro de experimentación

Siguiendo la lógica propuesta, encontramos dos puntos candidatos para el centro del primer diseño, que resultan simétricos con respecto al punto máximo. Podemos determinar asimismo las distancias relativas entre ambos puntos y el máximo \mathbf{x}^M dentro del plano de los factores, que las calculamos con el mismo MÓDULO 8, que nos lleva a comprobar que son prácticamente iguales, salvo errores de redondeo.

$$d_{A,M} = \sqrt{(x_1^A - x_1^M)^2 + (x_2^A - x_2^M)^2} = 1,3761$$

$$d_{B,M} = \sqrt{(x_1^B - x_1^M)^2 + (x_2^B - x_2^M)^2} = 1,3762$$

Debido a las condiciones de simetría, elegimos arbitrariamente uno de ellos, por ejemplo el \mathbf{B}^I , y lo tomamos como centro. Por tanto, el primer centro de experimentación será:

$$\mathbf{x}_0^I = (1,306447, -1,306447)'$$

A partir de este valor, surge de inmediato que la distancia que separa al máximo del primer centro estará dada por el mismo valor $d_{B,M}$, que de aquí en más identificaremos como d_{OM} . Así, este valor quedará como:

$$d_{OM} = 1,376237$$

En la fig. (6.3) puede verse claramente la representación de esta distancia y de las ubicaciones de todos los puntos que hemos definido en un gráfico en escala que hemos programado en el mismo MODULO 8, que se construye de modo automático teniendo como “inputs” solamente el modelo teórico —dado por el vector β elegido de los 60 disponibles— y el valor w seleccionado para definir la curva de nivel. En el **Apéndice C** pueden verse las líneas de comando correspondientes.

6.3. Primer factorial

6.3.1. Introducción

Situándonos en un contexto en donde hayamos definido la localización de un primer centro de diseño como hemos descrito anteriormente, exploraremos en un primer momento la superficie mediante un encadenamiento secuencial de diseños factoriales a dos niveles mientras valga la hipótesis de linealidad para los modelos ajustados. Definiremos a continuación los puntos que formarán este primer diseño, valiéndonos de algunos criterios propuestos los cuales discutiremos más adelante.

6.3.2. Lados del factorial y primer diseño

Partimos de la localización conocida del primer centro de experimentación, \mathbf{x}_0^I , que geoméricamente ubicaremos dentro del primer factorial como centro de un cuadrado de lado L . Ya que no contamos con criterios previos lo suficientemente informativos como para establecer con cierta eficiencia cuánto debería valer este lado, seguiremos una solución de compromiso que describimos a continuación.

Seguiremos a este efecto un criterio según el cual el valor que daremos al lado L del cuadrado será una fracción de la distancia que separa el centro de experimentación del máximo, d_{OM} . En otras palabras, tomaremos un valor de L comprendido entre 0 y d_{OM} , al que podemos traducir considerando un factor de proporcionalidad h , que varíe entre 0 y 1, de tal modo que el L elegido sea de la forma:

$$L = h \cdot d_{OM},$$

siendo:

$$d_{OM} = \|\mathbf{x}_0^I \mathbf{x}^M\|,$$

es decir, una fracción de la distancia entre el primer centro y el punto máximo de la ST .

De los puntos anteriores sabemos que: $d_{OM} = 1,376237$. Tomamos un valor arbitrario $h = 25\%$, con lo cual el valor de partida⁹ que consideraremos para el lado L del cuadrado¹⁰ será:

$$L = 0,25 \cdot 1,3762 = 0,3440593$$

Con este valor así determinado, podemos tomar entornos de amplitud igual a $\frac{1}{2}L$ con centro en \mathbf{x}_0^I y calcular las coordenadas de los puntos $\mathbf{C}^I, \mathbf{D}^I, \mathbf{E}^I$ y \mathbf{F}^I , que definen el cuadrado representativo del primer diseño, cuyo cálculo hemos sistematizado mediante las líneas de programa correspondientes al MÓDULO 2.1 de los programas de la serie CRJER:

```
> DF.01.01[, c('x1', 'x2')]
      x1      x2
C.1 1.20323 -1.40966
D.1 1.40966 -1.40966
E.1 1.40966 -1.20323
F.1 1.20323 -1.20323
O.1 1.30645 -1.30645
```

La representación gráfica de esta situación puede verse en la figura (6.5), y los comandos utilizados para representarla se encuentran al final de los programas de la serie MEDCA:

Y a partir de este *input*, calculamos los valores que toma la función $\pi(\mathbf{x}, \boldsymbol{\beta})$ para dichos puntos, que serán determinados mediante el proceso generador de datos, dado por la superficie teórica¹¹. Para ello, utilizamos la función `f.02`, del MÓDULO 7, programas de la serie BETAS. El aspecto que toma la tabla anterior es entonces el siguiente:

⁹En cuanto a la elección de qué fracción consideraremos como lado del factorial, proponemos partir de 5 niveles para el mismo, de modo que se corresponda cada uno de ellos con un conjunto de particiones de d_{OM} . Veremos más adelante que consideraremos como niveles de fraccionamiento de dicha distancia a los factores: 60%, 80%, 100%, 120% y 140%, los cuales evaluaremos para determinar cuál de ellos nos conduce a las mejores soluciones. Queda claro que la falta de información previa del proceso generador de datos nos impone proponer una solución de partida arbitraria. Sin embargo, también evaluaremos distintos valores para L , alrededor de un entorno del valor calculado para d_{OM} . En el capítulo siguiente, por ejemplo, mostraremos cómo evoluciona la estrategia al considerar 5 niveles de las variables L y S .

¹⁰En el comienzo de la experimentación, cuando carecemos de información relevante acerca del proceso y las condiciones geométricas del diseño, nos ha parecido que tomar lados del factorial iguales a d_{OM} resulte tal vez algo excesivo. Como para tener un valor inicial de referencia, elegimos entonces una fracción igual a la cuarta parte de esta distancia.

¹¹Por este motivo, diremos que estos puntos son también “teóricos” en el sentido en que fueron generados “sin ruido”, a partir de la superficie teórica.

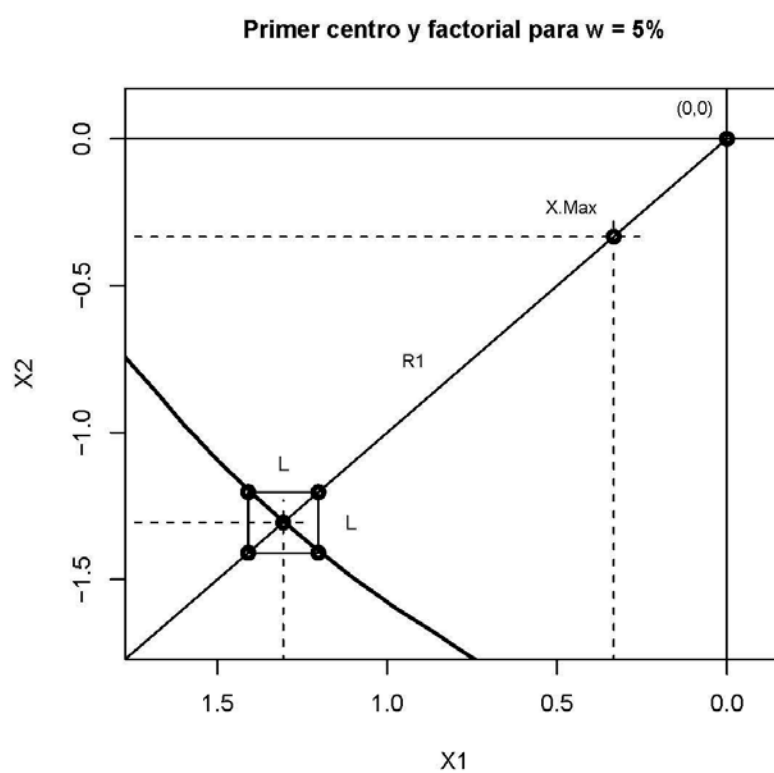


Figura 6.5: Situación del primer diseño en el plano de los factores: el centro de experimentación se determina al cortar la curva de nivel correspondiente a $\pi_5\%_{\text{máx}}$ con la recta que pasa por el origen y el máximo.

```
> DF.01.01[, c('x1', 'x2', 'Pi')]
      x1      x2      Pi
C.1 1.20323 -1.40966 0.04581
D.1 1.40966 -1.40966 0.01360
E.1 1.40966 -1.20323 0.04581
F.1 1.20323 -1.20323 0.13312
O.1 1.30645 -1.30645 0.04675
```

Nota bene: puede comprobarse que la función evaluada en el centro del diseño coincide con el resultado obtenido en la ecuación (6.3), ya que en la nomenclatura de \mathbf{R} tendremos:

```
> f.02(0.01[1], 0.01[2])
[1] 0.046751
```

que equivale, en efecto, a:

$$\pi(\mathbf{x}_0, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \mathbf{x}'_0 \mathbf{b} + \mathbf{x}'_0 \mathbf{B} \mathbf{x}_0)}{1 + \exp(\beta_0 + \mathbf{x}'_0 \mathbf{b} + \mathbf{x}'_0 \mathbf{B} \mathbf{x}_0)} = 0,046751$$

6.3.3. Generación de puntos simulados

El paso natural siguiente será el de obtener puntos simulados representativos del proceso “real”, no ya del “teórico”. De acuerdo con las consideraciones que hemos venido comentando, los puntos simulados tendrán las siguientes características:

- Para cada punto del primer diseño (5 en total), corresponderá un cierto número de éxitos, y_i , ($i = \text{C.01}, \text{D.01}, \text{E.01}, \text{F.01}$ y O.01), uno para cada punto, en principio.
- Cada número de éxitos provendrá de un proceso binomial de generación de datos, los que supondremos independientes entre sí, de parámetros m_i y $\pi(\mathbf{x}_i, \boldsymbol{\beta})$. Este último valor, de naturaleza teórica, es el valor que le corresponde a cada punto en estudio sobre la superficie teórica:

$$y_i \stackrel{\text{indep.}}{\sim} \text{binom.}[m_i, \pi(\mathbf{x}_i, \boldsymbol{\beta})], \quad 0 \leq y_i \leq m_i.$$

Desde el punto de vista estadístico, la probabilidad teórica $\pi(\mathbf{x}_i, \boldsymbol{\beta})$ será el valor esperado que tendrán los puntos generados, cuyos valores se encontrarán dispersos pero centrados en la probabilidad teórica. La fig. 6.6 permite visualizar esta situación.

- Fijando un número de repeticiones m_i del proceso de Bernoulli para cada punto —que define el llamado *index* o índice de la distribución [vid. p. ej.: MCCULLAGH Y NELDER (1989)]—, se calcularán las llamadas *proporciones estimadas* de cada uno de estos puntos, las cuales denotaremos como $\hat{\pi}(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{y_i}{m_i}$, para cada uno de los puntos del diseño.
- Para la simulación del número de éxitos correspondientes a cada punto del diseño, utilizaremos la función de generación de valores aleatorios para la distribución binomial que trae consigo el programa **R**:

```
> rbinom(n, size, prob)
```

La misma devuelve 'n' valores tomados al azar, cada uno de ellos provenientes de una distribución binomial de índice 'size' y probabilidad de éxito igual a 'prob'. Esta última corresponde al valor teórico de la función en el punto en cuestión, que además es el valor esperado de la distribución binomial. En cuanto al tamaño n de la muestra de éxitos y al índice size, y de acuerdo con esta estrategia planteada, no disponemos de información suficiente como para su asignación de modo eficiente. Por tanto, tomaremos inicialmente para el índice un valor de $m_i = 100$ como punto de partida, que también supondremos que será el mismo para cada uno de los puntos de diseño. La figura 6.6 nos ilustra esta situación para un punto genérico $\mathbf{x}_0 = (x_1, x_2)'$, cuyo valor sobre la superficie teórica hemos señalado como $\pi(\mathbf{x}_0, \boldsymbol{\beta})$ con un punto sin relleno. Los valores generados alrededor de este punto empleando la función `rbinom()` de **R** se han representado rellenos con fondo negro.

Lo que hemos realizado, lo repetimos para cada uno de los 4 puntos restantes que conforman el primer diseño, de manera que finalmente hayamos utilizado $n = 5m$ puntos en total del presupuesto disponible, que en nuestro caso serán $5 \times 100 = 500$ puntos. Correspondientemente, tendremos un vector de éxitos generados aleatoriamente de la forma $\mathbf{y} = (y_1, y_2, y_3, y_4, y_5)'$, en que la componente y_i representa el número de éxitos generados aleatoriamente para el punto de diseño i , es decir, para $i = \{\text{C.01}, \text{D.01}, \text{E.01}, \text{F.01}, \text{O.01}\}$.

A partir de esto, se puede definir el vector de proporciones observadas en cada punto de diseño dividiendo el número de éxitos de cada punto por el número m correspondiente a cada uno, que tomamos igual a 100:

$$\mathbf{pr} = \left(\frac{y_1}{100}, \frac{y_2}{100}, \frac{y_3}{100}, \frac{y_4}{100}, \frac{y_5}{100} \right)'$$

A los efectos de facilitar el cálculo de todas estas últimas cantidades, en los programas de la serie **CRJER** hemos desarrollado unas líneas de programa que permiten

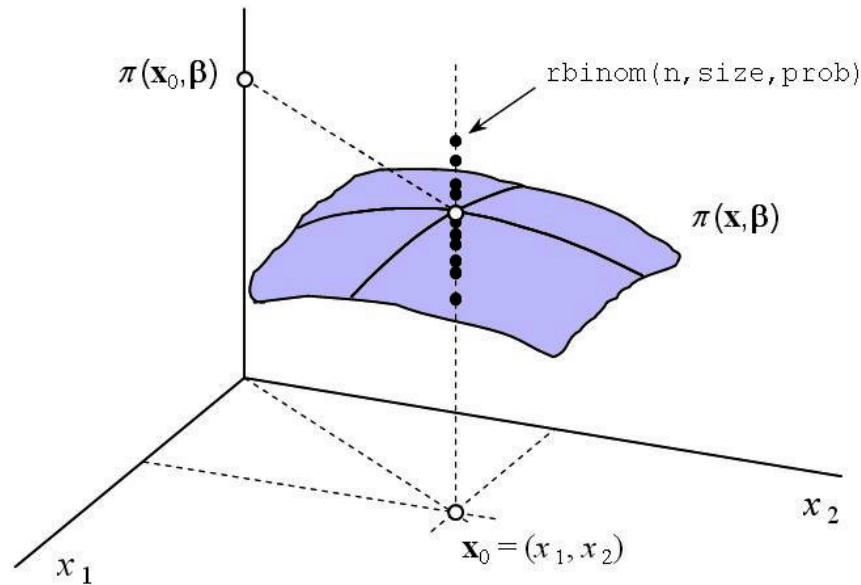


Figura 6.6: Generación aleatoria de puntos en torno al valor esperado de la función teórica, para un valor genérico de los factores. El valor $\pi(\mathbf{x}_0, \beta)$ será justamente el parámetro `prob` de esta función de **R** para este caso.

realizar esto de forma rápida y segura. Para el caso evaluado, la tabla de resultados que se obtiene es la siguiente, en la que `x1` y `x2` son las coordenadas de los puntos de diseño; `Pi` es el valor de dichos puntos sobre la superficie teórica; `M` es el número de repeticiones del experimento de Bernoulli con probabilidad `Pi`; `y` es el número de éxitos simulado y `pr` es la proporción observada sobre la base de la simulación:

```
> DF.01.01
      x1      x2      Pi  M  y  pr
C.1 1.20323 -1.40966 0.04581 100  5 0.05
D.1 1.40966 -1.40966 0.01360 100  2 0.02
E.1 1.40966 -1.20323 0.04581 100  6 0.06
F.1 1.20323 -1.20323 0.13312 100 13 0.13
O.1 1.30645 -1.30645 0.04675 100  7 0.07
```

6.4. Ajuste del primer modelo y primer salto

Teniendo ya 5 puntos de un primer diseño —ver la columna encabezada con la etiqueta `pr` de la salida anterior, `DF.01.01`— podemos ensayar un primer ajuste de

acuerdo a la naturaleza de los datos. Puesto que son puntos provenientes de procesos binomiales, en el contexto de los *MLG* tendremos que elegir una función “link” que vincule el valor esperado de la respuesta con un predictor lineal apropiado.

6.4.1. El predictor lineal

Puesto que disponemos de $n = 5$ puntos para el primer diseño, el predictor lineal que tomemos en consideración podrá definirse a lo sumo con $p = 5$ parámetros por razones estructurales comúnmente conocidas. Por otra parte, y ya que no sería conveniente considerar modelos de ajuste de mayor grado de complejidad que el modelo teórico, dispondremos de un máximo de 6 parámetros para elegir, los mismos 6 que corresponden al modelo teórico.

Consideraremos para cada punto de diseño, pues, un predictor lineal de 5 términos de la forma:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \beta_{12} x_{i1} x_{i2} + \beta_{11} x_{i1}^2,$$

cuyos coeficientes debemos estimar.

6.4.2. El modelo logístico

Por los motivos que hemos explicado anteriormente —que pueden complementarse con lo indicado en el **Apéndice B**— el modelo logístico resulta un modelo adecuado para ajustar datos provenientes de procesos binomiales. Será entonces éste el que utilizaremos para ajustar funciones de aproximación de la superficie teórica en lo sucesivo.

La expresión genérica de este modelo al considerar $p = 5$ parámetros, será de la forma:

$$\pi(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \beta_{12} x_{i1} x_{i2} + \beta_{11} x_{i1}^2)}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \beta_{12} x_{i1} x_{i2} + \beta_{11} x_{i1}^2)}$$

El modelo estimado será de la misma forma, en la que aparecerán cada uno de los coeficientes originales β_j reemplazados por sus estimaciones máximo verosímiles $\hat{\beta}_j$.

Puesto que el argumento de la transformación logit lo podemos considerar como la proporción observada en cada condición experimental, tendremos que $\pi(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{pr}_i$, con lo cual nos encontramos en condiciones de estimar los parámetros del modelo. La función que permite realizar esto en **R** es `glm()`, definiendo adecuadamente sus parámetros.

Un primer ajuste para estos puntos utilizando la función precitada se resume a continuación, en donde las primeras dos columnas (**x1** y **x2**) representan las coordenadas en el plano de los factores de los 5 puntos considerados para el primer diseño,

que están representados por cada una de las filas. Los valores de la probabilidad de éxito para cada punto en la superficie teórica, el número de repeticiones de Bernoulli para cada punto y la proporción observada para cada uno, se expresan en las últimas 4 columnas, encabezadas con los nombres: P_i , M , y y pr , respectivamente:

```
> DF.01.01
      x1      x2      Pi  M  y  pr
C.1 1.20323 -1.40966 0.04581 100  5 0.05
D.1 1.40966 -1.40966 0.01360 100  2 0.02
E.1 1.40966 -1.20323 0.04581 100  6 0.06
F.1 1.20323 -1.20323 0.13312 100 13 0.13
O.1 1.30645 -1.30645 0.04675 100  7 0.07

> glm.1 <- glm(pr ~x1 + x2 + x12 + x11, family = binomial, weights = m)
```

El resultado de las estimaciones de los coeficientes resulta, pues:

```
> summary.glm(glm.1)$coefficients
      Estimate Std. Error  z value Pr(>|z|)
(Intercept) -39.612558   85.62537 -0.46262639 0.6436322
x1           68.647691  120.38297  0.57024419 0.5685121
x2            2.321508   29.58489  0.07846937 0.9374547
x12           2.271701   23.30733  0.09746723 0.9223554
x11          -26.803549   43.55162 -0.61544317 0.5382621
```

6.4.3. Selección de términos del modelo ajustado

Generalidades

Dos de los criterios habitualmente empleados¹² para seleccionar los términos del modelo ajustado son:

- a. *El criterio del “p-value” crítico*, que considera como significativos aquellos términos cuyos parámetros estimados tengan un “p-value” inferior a un cierto valor crítico, tomado habitualmente igual al 5%. Esto se repite para cada uno de los coeficientes, y quedarán en el modelo final los que satisfagan esta condición.

¹²En WU Y HAMADA (2000) y en BURNHAM Y ANDERSON (2002), por ejemplo, pueden verse otros criterios.

- b. *El criterio jerárquico*, que establece que si el término que contiene el mayor orden de la variable resulta significativo, entonces el resto de los términos de la misma variable y de menor orden también se considerarán parte del modelo ajustado final. Del mismo modo, este criterio constituye una valiosa herramienta para comparar dos modelos y conducir el proceso de selección del modelo final [vid. LOSILLA *et al.* (2005), pp. 206 *et seq.*; COLLETT (2003), p. 74, entre otros].

Breve discusión: criterio del “p-value” versus criterio jerárquico

Una pregunta razonable con la que comenzar la cuestión de decidir qué criterio adoptar para la selección de términos podría ser: *¿tiene sentido considerar en el modelo final aquellos términos cuyos coeficientes sean no significativos, siendo que los términos de orden superior en la misma variable sí lo sean?* Para intentar llegar a una posición objetiva, consideraremos un ejemplo sencillo.

Sean M_1 y M_2 dos modelos candidatos estimados de una variable, con términos cuadráticos, tales que:

$$\begin{aligned} M_1 & : \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_{11} x^2 \\ M_2 & : \hat{y} = \hat{\beta}_0 + \hat{\beta}_{11} x^2 \end{aligned}$$

Supongamos también que, siguiendo algún criterio específico, hayamos determinado la significatividad de los coeficientes estimados, resumiéndose en la siguiente tabla:

COEF.	SIGNIF.
β_0	significativo
β_1	no significativo
β_{11}	significativo

Nos preguntamos a continuación: *¿con qué modelo nos quedamos: con M_1 o con M_2 ?*

La simple observación de la tabla anterior, pareciera sugerirnos que el segundo modelo fuera más adecuado que el primero, puesto que considerar el término lineal equivaldría a considerar un término cuyo coeficiente resulta no significativo. Sin embargo, y partiendo de la base según la cual el término cuadrático tiene un coeficiente que es significativo, veremos que eliminar el término lineal del modelo sería imponer restricciones “fuertes” que de antemano no conocemos de un fenómeno, también no conocido al principio, del que apenas hemos podido conseguir unas observaciones y ajustar un modelo estadístico que suponemos que es adecuado. A continuación, daremos algunas razones por las que creemos que no perdemos nada en dejar el término lineal en el modelo final y que también ganamos en flexibilidad del modelo ajustado al fenómeno estudiado.

Razones estructurales

Resulta obvio que la representación de ambos modelos refiere a una función parabólica en una variable. Tomando, por ejemplo, el modelo M_1 , la abscisa de su punto crítico, $x = x_C$, estará dada por la solución en x de la siguiente ecuación:

$$\left(\frac{\partial \hat{y}}{\partial x}\right)_{x=x_C} = \frac{\partial}{\partial x} (\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_{11} x^2)_{x=x_C} = 0$$

De inmediato, encontramos que:

$$x_C = \frac{-\hat{\beta}_1}{2\hat{\beta}_{11}}, \quad \text{con: } \hat{\beta}_{11} \neq 0$$

Para el numerador también distinto de cero, la parábola tendrá su vértice en el punto $x = x_C$ dado por la ecuación anterior, mientras que si el mismo numerador fuese igual a cero, lo anterior se traduce en que el vértice de la parábola se encontrará sobre la recta $x = 0$. En otras palabras: no rechazar la hipótesis nula según la cual $H_0 : \beta_1 = 0$, equivale a no rechazar que el punto crítico de la función cuadrática se encuentre centrado en el eje y . Si, en cambio, rechazáramos esta hipótesis en favor de $H_1 : \beta_1 \neq 0$, estaríamos conjeturando que la misma parábola se encontraría no centrada en el mismo eje. El criterio jerárquico de selección de términos dice que si $\hat{\beta}_{11}$ resulta significativo, entonces es razonable incluir también el término lineal en el modelo, cuyo coeficiente es $\hat{\beta}_1$, independientemente que éste resulte significativo o no.

Parece razonable pensar que, siguiendo el criterio clásico del “p-value” crítico — según el cual estaríamos dejando fuera del modelo al término $\hat{\beta}_1 x$ —, estaríamos forzando a que el modelo ajustado deba estar centrado en el eje y sin más información que la que nos han proporcionado los puntos de diseño. Aunque también parece razonable afirmar que imponerle esta restricción al sistema o fenómeno estudiado podría ser algo relativamente temerario, puesto que conocemos muy poco sobre su comportamiento.

En conclusión, pensamos que la incorporación del término lineal al modelo ajustado final no nos agrega ninguna dificultad en nuestro estudio. Antes bien, su consideración hace que el modelo ajustado sea más flexible y menos restrictivo a la hora de seguir aprendiendo cómo se relacionan funcionalmente las variables explicada y explicativa, sin estar imponiendo restricciones de las que no tenemos evidencia suficiente de que sean válidas. Razonamos también en esta misma línea y hacemos una extrapolación: para poder tener evidencias acerca de la conveniencia de considerar restricciones, sería necesario contar con más información que la obtenida a partir de un modelo estadístico.

Razones comparativas

En sintonía con el punto anterior, también podemos suponer el caso en que dos analistas, partiendo de la misma tabla de datos, se proponen ajustar modelos a los

mismos, cada uno con su punto de vista y su metodología. Inicialmente, podemos conjeturar también que existen razones para suponer que es válida la hipótesis de normalidad que sigue la variable respuesta de la tabla de datos, como así también la de independencia. Diremos además, que el diseño que se utilizó para recoger las observaciones de la respuesta es de naturaleza ortogonal.

El carácter ortogonal de la matriz de diseño, como bien se sabe, conduce a que las estimaciones de los parámetros del modelo a ajustar resulten independientes entre sí. Una consecuencia inmediata de esto es que la supresión de un término cualquiera del modelo no implica que haya que recalcular los coeficientes restantes, puesto que las estimaciones de todos ellos son independientes entre sí y dependen únicamente de la matriz de diseño, que es única.

Siguiendo con nuestro ejemplo, digamos que el primer analista ajusta un modelo normal en el que sigue el criterio clásico del “p-value” crítico, según el cual su modelo final sería: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_{11}x^2$. El segundo analista, en cambio, siguiendo el criterio jerárquico, decide incorporar el término lineal, con lo cual su modelo final sería: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1x + \hat{\beta}_{11}x^2$. Ambos analistas, además, completan sus respectivos informes con la validación correspondiente de las hipótesis, llegando a conclusiones similares.

¿Pero qué ocurriría si ambos decidiesen comparar sus resultados, es decir, sus modelos ajustados? Uno de ellos diría que al otro le sobra un término en su modelo, y el otro, que el uno ha suprimido un término innecesariamente. Sin embargo, hay un hecho que resulta contundente: si una tercera persona tuviese ambos modelos ajustados como única información, la tabla de estimaciones del modelo con $p = 3$ parámetros le resultará más adecuada que la que tiene $p = 2$. La diferencia entre uno y otro modelos radica en que a partir del modelo de $p = 3$ sería posible anular la estimación de β_1 y llegar a la misma tabla de estimaciones que la del modelo $p = 2$, mientras que el sentido inverso no es posible de realizar.

Concluyendo, diremos que nos parece más razonable presentar los datos del “modelo completo” frente a los del “modelo reducido”, puesto que la homologación de resultados podría ser menos restrictiva que si se presentaran conclusiones sobre la base de un modelo sujeto a condiciones. El analista que lo deseara, podría en su caso reducir fácilmente el modelo completo a otros más simples, aunque no sería posible ampliar un modelo reducido con la misma facilidad.

Reordenando conceptos, diremos que no hay un *único* e infalible método de selección de términos en los modelos ajustados. De este modo, y por las razones que hemos expuesto, en lo que sigue de nuestro trabajo nos inclinaremos a considerar los modelos ajustados cuyos términos se hayan seleccionado a partir del criterio jerárquico, ya que

nos parece más flexible que el criterio del “p-value” crítico. Mientras no especifiquemos lo contrario, seguiremos entonces el criterio jerárquico.

Otras discusiones acerca de la aplicabilidad de este modelo pueden encontrarse en la referencia precitada, la cual aborda también otras referencias adicionales.

6.4.4. El modelo elegido

Ajustaremos a continuación un modelo mediante selección “hacia atrás”, partiendo del modelo con mayor número de parámetros que podemos considerar ($p = 5$) y siguiendo el criterio jerárquico para suprimir términos. Consideraremos un “valor de salida” arbitrario¹³ igual al 15 %, algo más permisivo que el “clásico 5 %”. Así, y volviendo a examinar nuestro primer modelo ajustado, encontramos que el término de mayor orden, $\hat{\beta}_{11}x^2$, tiene su coeficiente no significativo. Esto nos lleva a considerar un segundo modelo ajustado en el que suprimimos el término cuadrático, ya que su “p-value” resulta superior al valor de corte del 15 % que hemos fijado de antemano:

```
> glm.2 <- glm(pr ~ x1 + x2 + x12, family = binomial, weights = m)
```

```
> summary.glm(glm.2)$coefficients
              Estimate Std. Error   z value Pr(>|z|)
(Intercept)  21.044481   40.39599  0.5209548 0.6023983
x1           -10.401183   30.91386 -0.3364569 0.7365264
x2            17.602987   32.46399  0.5422313 0.5876592
x12           -7.378543   24.83179 -0.2971410 0.7663589
```

Siguiendo la misma metodología, suprimimos el término de la interacción y reajustamos el modelo:

```
> glm.3 <- glm(pr ~ x1 + x2, family = binomial, weights = m)
```

```
> summary.glm(glm.3)$coefficients
              Estimate Std. Error   z value Pr(>|z|)
(Intercept)  9.134808   3.998599  2.2845020 0.0223420474
x1           -1.236175   2.004387 -0.6167347 0.5374097456
x2            8.013523   2.431243  3.2960607 0.0009805083
```

¹³Algunos paquetes estadísticos de uso corriente en estadística industrial, como por ejemplo el Minitab, consideran este mismo valor del 15 % por defecto. En BURNHAM Y ANDERSON (2002), p. 36, por ejemplo, los autores citan a RAWLINGS *et al.* (1988), en donde se discute brevemente el valor considerado y se dan razones de la elección de este mismo valor en ajustes del tipo “stepwise regression”.

Y repetimos lo propio con el factor x_1 :

```
> glm.4 <- glm(pr ~ x2, family = binomial, weights = m)

> summary.glm(glm.4)$coefficients
              Estimate Std. Error z value Pr(>|z|)
(Intercept) 7.530384    3.052353 2.467075 0.0136221664
x2           8.017551    2.432242 3.296362 0.0009794571
```

De esta manera, llegamos a que el `glm.4` parece un modelo razonable para ajustar los datos del primer factorial, con lo que el modelo ajustado quedará como:

$$\hat{\pi}(\mathbf{x}, \hat{\boldsymbol{\beta}}) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)} = \frac{\exp(7,53 + 8,02x_2)}{1 + \exp(7,53 + 8,02x_2)}$$

Dado que el programa **R** no trae incorporada una macro que permita la selección de términos de acuerdo con el criterio jerárquico, hemos programado una serie de comandos que podrán encontrarse en la serie de programas **CRJER** del **Apéndice C**, en donde a partir de un conjunto de puntos se ajustan de forma sucesiva una serie de modelos que siguen este criterio jerárquico, de modo de llegar de forma sistemática y automática al modelo ajustado. En el **MÓDULO 2.2** de los mismos hemos propuesto los comandos necesarios para ajustar modelos con 5 puntos de diseño, como es el caso del primer factorial. En el **MÓDULO 3.2**, por ejemplo, reformulamos el mismo programa para ajustar modelos con 6 ó más puntos de diseño¹⁴.

6.4.5. El primer salto

Así como comentamos el problema con que nos encontramos a la hora de definir valores adecuados para L , con S nos ocurre algo similar: no tenemos información alguna acerca de qué valor considerar como adecuado (ver anteriormente la sección “Lados del factorial y primer diseño”, en donde presentamos algunos motivos de la elección de valores de compromiso debido a esta falta de información). Dentro de una gama propuesta de valores que comentaremos más adelante —cuando abordemos la definición de la matriz **LS**—, partiremos de un valor inicial arbitrario para S que sea igual al 20% de la misma distancia que tenemos de referencia, d_{OM} . Este valor de partida, 20% de d_{OM} , implica considerar módulos de salto equivalentes a la quinta

¹⁴Tanto en una rutina como en la otra, y a la hora de haber programado las mismas, hemos comparado las estimaciones obtenidas con las que obtuvimos de los mismos ajustes realizados “manualmente” o “paso a paso”, llegando a los mismos resultados.

parte de aquella distancia, cifra que consideramos adecuada como punto de partida en el escenario de falta de información adicional.

Para determinar la dirección que tendrá este primer salto, nos valemos del criterio del “steepest path” o camino de máxima pendiente de crecimiento del modelo ajustado, herramienta en la que se basa el “modelo clásico”. Según esto, se seguirá aquella dirección definida por el vector cuyas componentes sean:

$$\mathbf{d} = \left(\frac{\partial}{\partial x_1} [\hat{\pi}(\mathbf{x}, \hat{\boldsymbol{\beta}})], \frac{\partial}{\partial x_2} [\hat{\pi}(\mathbf{x}, \hat{\boldsymbol{\beta}})] \right)'$$

Para el modelo genérico ajustado:

$$\hat{\pi}(\mathbf{x}, \hat{\boldsymbol{\beta}}) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_1 x_2 + \hat{\beta}_{12} x_1 x_2 + \hat{\beta}_{11} x_1^2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_1 x_2 + \hat{\beta}_{12} x_1 x_2 + \hat{\beta}_{11} x_1^2)},$$

tendremos que las derivadas parciales de $\hat{\pi}(\mathbf{x}, \hat{\boldsymbol{\beta}})$ con respecto a los dos factores son análogas a las derivadas parciales correspondientes del predictor lineal del modelo:

$$\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_1 x_2 + \hat{\beta}_{12} x_1 x_2 + \hat{\beta}_{11} x_1^2$$

Por lo tanto, consideraremos las siguientes expresiones¹⁵:

$$\begin{aligned} \frac{\partial \hat{\eta}}{\partial x_1} &= \hat{\beta}_1 + \hat{\beta}_{12} x_2 + 2\hat{\beta}_{11} x_1 \\ \frac{\partial \hat{\eta}}{\partial x_2} &= \hat{\beta}_1 + \hat{\beta}_{12} x_1 \end{aligned}$$

Una vez obtenidos estos valores, el valor absoluto del ángulo que define la dirección del vector S se calculará como:

$$\arctan \left[\frac{(\partial \hat{\eta} / \partial x_2)}{(\partial \hat{\eta} / \partial x_1)} \right],$$

del cual se determina el cuadrante correspondiente mediante los signos de las dos componentes. Estos cálculos los hemos implementado de forma automática en módulos como el 2.3.01 de los programas de la serie **CRJER**, que devuelve el valor final del ángulo φ expresado en grados sexagesimales y las componentes del salto sobre ambos ejes:

$$\begin{aligned} s_1 &= S \cdot \cos \varphi \\ s_2 &= S \cdot \sen \varphi \end{aligned}$$

Para nuestro caso de estudio, hemos calculado ambas derivadas con **R** teniendo en cuenta todos los coeficientes del modelo completo, llegándose a los valores del ángulo φ_1 , componente de S sobre x_1 y componente sobre x_2 , respectivamente, que quedan representados mediante los valores **phi5.g.01** y **S5.01** como salidas de **R**:

¹⁵Cuando ambas expresiones dependiesen de x_1 y/o de x_2 , ambas derivadas se evaluarán en las coordenadas respectivas del centro del diseño que sea origen del salto. Para el caso del primer diseño, se tomarían los puntos $x_1 = x_{01}$ y $x_2 = x_{02}$.

```
> phi5.g.01
[1] 95.27759
```

```
> S5.01
[1] -0.02531755  0.27408057
```

6.5. Segundo factorial

Un problema que podría presentarse a la hora de considerar el modelo ajustado es que el único término que forme parte del mismo sea el intercepto o término independiente. En esos casos, no sería posible determinar pendiente alguna de máximo crecimiento por carecer de los valores tanto de una como de la otra derivada del predictor lineal.

Para protegernos de estos casos, hemos considerado lo siguiente: *si no puede calcularse la dirección de máximo crecimiento de la respuesta, se establece que el salto será igual a cero y el segundo diseño se toma haciendo coincidir el primer centro con el segundo, y rotando 45° los puntos periféricos del diseño*¹⁶. En los otros casos, en donde pueda determinarse la dirección del salto, los puntos del segundo diseño se obtendrán a partir de los del primero, sumándole a cada uno el vector salto obtenido. Esquemáticamente, se tiene la siguiente ecuación vectorial:

$$\mathbf{P}_2 = \mathbf{P}_1 + \mathbf{S}_{12}$$

En módulos como por ejemplo el 3.1.1 de los programas de la serie CRJER hemos programado los comandos necesarios para que se tengan en cuenta estas dos situaciones, y que el programa pueda determinar si es necesario o no obtener el segundo diseño sumándole el vector salto o bien rotándolo alrededor del centro. Con los valores del salto y del ángulo calculados en el punto anterior, el programa nos permite determinar los puntos del segundo diseño.

Una vez que tenemos los 5 puntos del segundo diseño, repetimos la misma rutina de cálculos que hicimos con el primer diseño:

1. Cálculo de la probabilidad teórica para los puntos del segundo diseño.
2. Generación aleatoria de éxitos para cada punto, con probabilidad de éxito igual a la calculada en el punto 1.

¹⁶Considerar un diseño rotado con respecto a uno que tenga los lados paralelos a los ejes deja de ser un diseño factorial en el *sentido estricto* del término.

3. Determinación de la proporción observada de éxitos.

El mismo programa nos permite hacer automáticamente estos cálculos y nos presenta el estado de situación de los primeros dos diseños:

```
> DF.02.01
          x1          x2          Pi  M  y  pr
C.1 1.203230 -1.4096600 0.045810 100  2 0.02
D.1 1.409660 -1.4096600 0.013600 100  2 0.02
E.1 1.409660 -1.2032300 0.045810 100  9 0.09
F.1 1.203230 -1.2032300 0.133120 100 12 0.12
O.1 1.306450 -1.3064500 0.046750 100  6 0.06
C.2 1.177911 -1.1355838 0.197452 100 11 0.11
D.2 1.384347 -1.1355838 0.074733 100  6 0.06
E.2 1.384347 -0.9291482 0.181681 100 19 0.19
F.2 1.177911 -0.9291482 0.383109 100 38 0.38
O.2 1.281129 -1.0323660 0.192734 100 27 0.27
```

Obsérvese que las primeras 5 filas corresponden a los puntos del primer diseño, y las segundas, a los del segundo diseño. También puede verse que la región de experimentación ha “migrado” de la región de vértices C.1-D.1-E.1-F.1 a la definida por C.2-D.2-E.2-F.2, cada una con centro en O.1 y O.2, respectivamente. En el mismo programa de la serie CRJER podrán encontrarse unas líneas en R que permitirán visualizar la situación de los puntos experimentales anteriores en el plano de los factores, cuyo resultado podemos ver en la figura siguiente (fig. 6.7). Nótese que el salto —en dirección y sentido— hace que los puntos del segundo diseño se vayan acercando más hacia el máximo que lo que lo hacía el primer diseño:

Hasta el momento, con los puntos observados en este segundo diseño, ya se llevan utilizados 1000 puntos de los 1500 disponibles que nos habíamos propuesto emplear.

6.6. Ajuste del segundo modelo y segundo salto

Contando entonces con los 10 puntos resultantes del primero y segundo diseño, podemos ajustar un nuevo modelo de las mismas características que el ajustado en el primer diseño, con la salvedad que en este caso podemos tomar el predictor lineal completo de segundo orden, $p = 6$, del que probablemente podremos obtener más información:

$$\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1 x_2 + \hat{\beta}_4 x_1^2 + \hat{\beta}_5 x_2^2$$

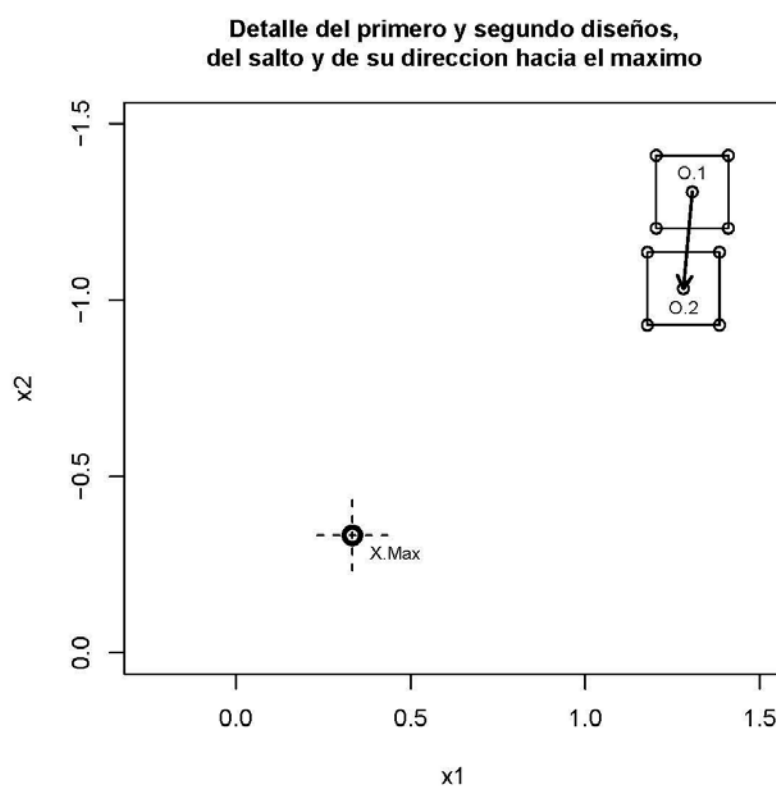


Figura 6.7: Primero y segundo diseños y su aproximación hacia el punto máximo.

Empleando el ajuste automático de acuerdo con el criterio jerárquico que puede verse, por ejemplo, en el MÓDULO 3.2.01 de la serie de programas CRJER, llegamos al siguiente modelo ajustado, que coincide con el que se hubiera ajustado de forma manual:

```
> DF6.01
              Estimate Std. Error  z value  Pr(>|z|)
(Intercept) -43.66063   32.837269 -1.329606 0.18364811
x.01          94.02248   52.122283  1.803883 0.07124972
x.02          24.63574   10.114572  2.435668 0.01486433
x.12         -14.80031    7.828236 -1.890632 0.05867353
x.11         -43.86917   21.035291 -2.085503 0.03702365
x.22           0.00000    0.000000  0.000000 0.00000000
```

Siguiendo la misma metodología que cuando pasamos del primero al segundo diseño, en el MÓDULO 3.3 de los programas de la serie CRJER calculamos el valor del ángulo ($\phi_{6.g.01}$) y la dirección del segundo salto (vector $S_{6.01}$) que unirá el segundo con el tercer diseño. Los resultados son los siguientes:

```
> phi6.g.01
[1] 95.85746

> S6.01
[1] -0.0280901  0.2738103
```

6.7. Tercer factorial

Teniendo ya el segundo salto y su dirección, podemos calcular la ubicación de los puntos del tercer factorial siguiendo la misma lógica que para el segundo y el primero. Asimismo, mediante los mismos programas podemos calcular las probabilidades de dichos puntos sobre la superficie teórica, el número de éxitos generados para cada punto y las proporciones observadas, respectivamente. (Gráfico de referencia: fig. 6.8). En el MÓDULO 4.1 de la serie CRJER aparecen las líneas de programa que utilizamos para estos cálculos, cuyo resultado final es:


```

> DF.03.01
      x1      x2      Pi  M  y  pr
C.1 1.203230 -1.4096600 0.045810 100  2 0.02
D.1 1.409660 -1.4096600 0.013600 100  2 0.02
E.1 1.409660 -1.2032300 0.045810 100  9 0.09
F.1 1.203230 -1.2032300 0.133120 100 12 0.12
O.1 1.306450 -1.3064500 0.046750 100  6 0.06
C.2 1.177911 -1.1355838 0.197452 100 11 0.11
D.2 1.384347 -1.1355838 0.074733 100  6 0.06
E.2 1.384347 -0.9291482 0.181681 100 19 0.19
F.2 1.177911 -0.9291482 0.383109 100 38 0.38
O.2 1.281129 -1.0323660 0.192734 100 27 0.27
C.3 1.149821 -0.8617735 0.477998 100 49 0.49
D.3 1.356257 -0.8617735 0.256225 100 34 0.34
E.3 1.356257 -0.6553379 0.427460 100 44 0.44
F.3 1.149821 -0.6553379 0.645694 100 66 0.66
O.3 1.253039 -0.7585557 0.457891 100 46 0.46

```

Podemos comprobar que habiendo llegado al tercer diseño, ya hemos utilizado todo el presupuesto disponible que nos habíamos propuesto, que era igual a 1500 puntos de diseño. Los dos pasos inmediatos que siguen a esto serán los de ajustar el modelo final con los 1500 puntos y evaluar la calidad del ajuste conseguido. En los puntos siguientes desarrollaremos estos aspectos.

6.8. Ajuste del modelo final

Contando entonces con los 15 puntos de diseño —o, dicho de otro modo, con los 1500 experimentos de Bernoulli— podemos ajustar un último modelo, que será con el que finalmente evaluaremos la calidad de las estrategias que elegimos. En el MÓDULO 4.2.01 de la serie de programas CRJER pueden verse las líneas de programa escritas para realizar este ajuste automáticamente, siguiendo el criterio jerárquico. El resultado final es:

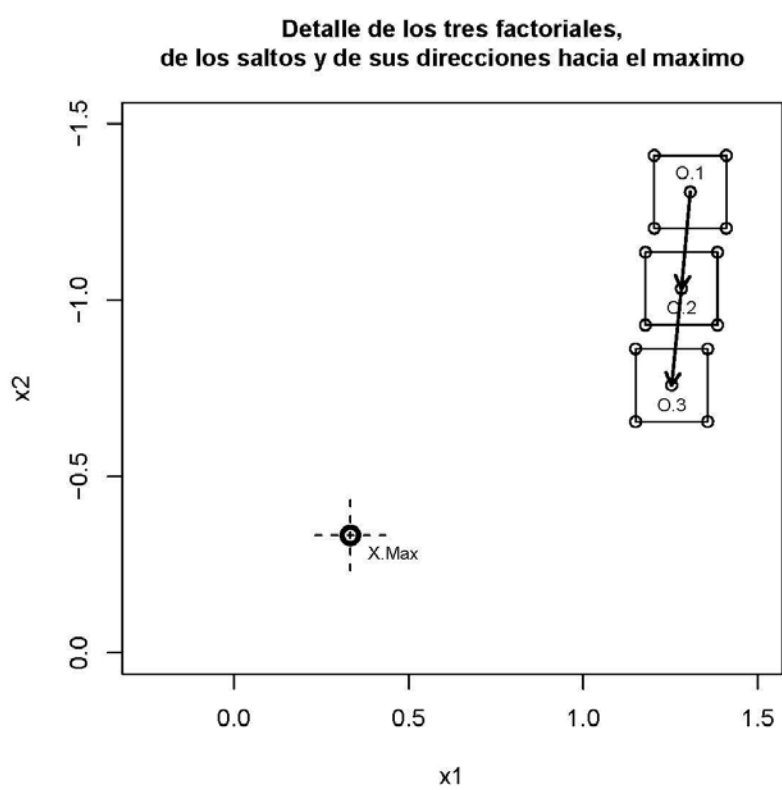


Figura 6.8: Los tres factoriales aproximándose secuencialmente hacia el máximo.

```
> DF7.01
              Estimate Std. Error   z value   Pr(>|z|)
(Intercept)  9.750339  0.6450672  15.115229 1.285239e-51
x.01         -4.595659  0.4332252 -10.608014 2.735007e-26
x.02          5.137089  0.6567979   7.821416 5.223249e-15
x.12          0.000000  0.0000000   0.000000 0.000000e+00
x.11          0.000000  0.0000000   0.000000 0.000000e+00
x.22          0.000000  0.0000000   0.000000 0.000000e+00
```

De esta manera, vemos que el modelo que mejor ha ajustado los 15 puntos de diseño ha sido un modelo lineal en los factores, sin interacción ni términos cuadráticos.

6.9. El próximo paso

Las secciones que describimos a lo largo de este capítulo nos llevaron a generar una superficie de respuesta con la cual pudimos generar aleatoriamente un conjunto de 3 factoriales sucesivos, cada uno consistente en 5 puntos de diseño los que, a su vez, condensaban la información de 100 repeticiones de experimentos de Bernoulli. Con ello, logramos disponer de un cierto modo de toda la información que estaba contenida en los 1500 puntos en total, valor que alcanzaba nuestro presupuesto fijo inicialmente establecido. Con todos estos puntos, llegamos a un modelo final ajustado que no nos dice mucho por sí solo, dado que no podemos evaluar comparativamente si el ajuste alcanzado resultó ser “relativamente bueno” o “relativamente malo” en comparación con otros.

En el capítulo siguiente, concentraremos nuestros esfuerzos en generar varios casos de puntos simulados, de modo que podamos realizar una comparación relativa de los ajustes conseguidos. A tal efecto, generaremos unos 15 conjuntos de datos a partir de los mismos valores que tomamos para los parámetros de diseño: w , L y S que consideramos en el presente capítulo, y comenzaremos a preparar el terreno para hacer las comparaciones de los ajustes conseguidos en cada caso de acuerdo con los dos criterios que hemos propuesto: el de *cantidad de información* y el de *proximidad al máximo*. Habiendo conseguido, pues, desarrollar varios casos para las mismas variables de estudio y compararlos entre sí, en los capítulos 8 y 9 estaremos ya en condiciones de explotar la consideración de múltiples niveles para ellos, que nos permitirá evaluar los ajustes de forma más amplia y explorar la posibilidad de buscar condiciones óptimas basadas en los criterios que propusimos.

El apoyo de ciertos programas que hemos escrito en **R** se nos hace imprescindible para llevar a cabo este paso de modo exitoso. En el **Apéndice C** reproducimos las

principales líneas de programa utilizadas.

Nota: en adelante, nos referiremos a cada conjunto de 1500 puntos de diseño generados por un mismo valor de variables de estudio w , L y S como un *cuadro de puntos*, para distinguirlo de lo que más adelante llamaremos *caso de estudio*¹⁷, que serán los distintos *cuadros de puntos* evaluados para cada uno de los niveles considerados para esta terna de variables de estudio. De este modo, si consideramos 5 niveles para cada una de estas variables de estudio, esto se traducirá en el estudio de $5 \times 5 \times 5 = 125$ casos de estudio, cada uno compuesto por 15 cuadros de puntos, los que a su vez, estarán formados por 1500 realizaciones de la respuesta binaria “éxito-fracaso”, contabilizando en nuestro caso los éxitos. El siguiente esquema gráfico (figura 6.9) ilustra las ideas precisadas para el estudio de 1 cuadro de puntos:

¹⁷En la nomenclatura empleada por el programa **R**, diremos que en el capítulo anterior hemos analizado un “data frame” de puntos, consistente en 15 puntos de diseño y un total de 1500 observaciones realizadas de la respuesta binaria “éxito-fracaso”, que luego tradujimos en términos de proporciones, a las que finalmente ajustamos modelos logísticos siguiendo el criterio jerárquico de selección de términos. En adelante, la referencia **DF** se referirá a esta concepción de “data frame” o “cuadro de puntos” que hemos dado al conjunto de puntos estudiados.

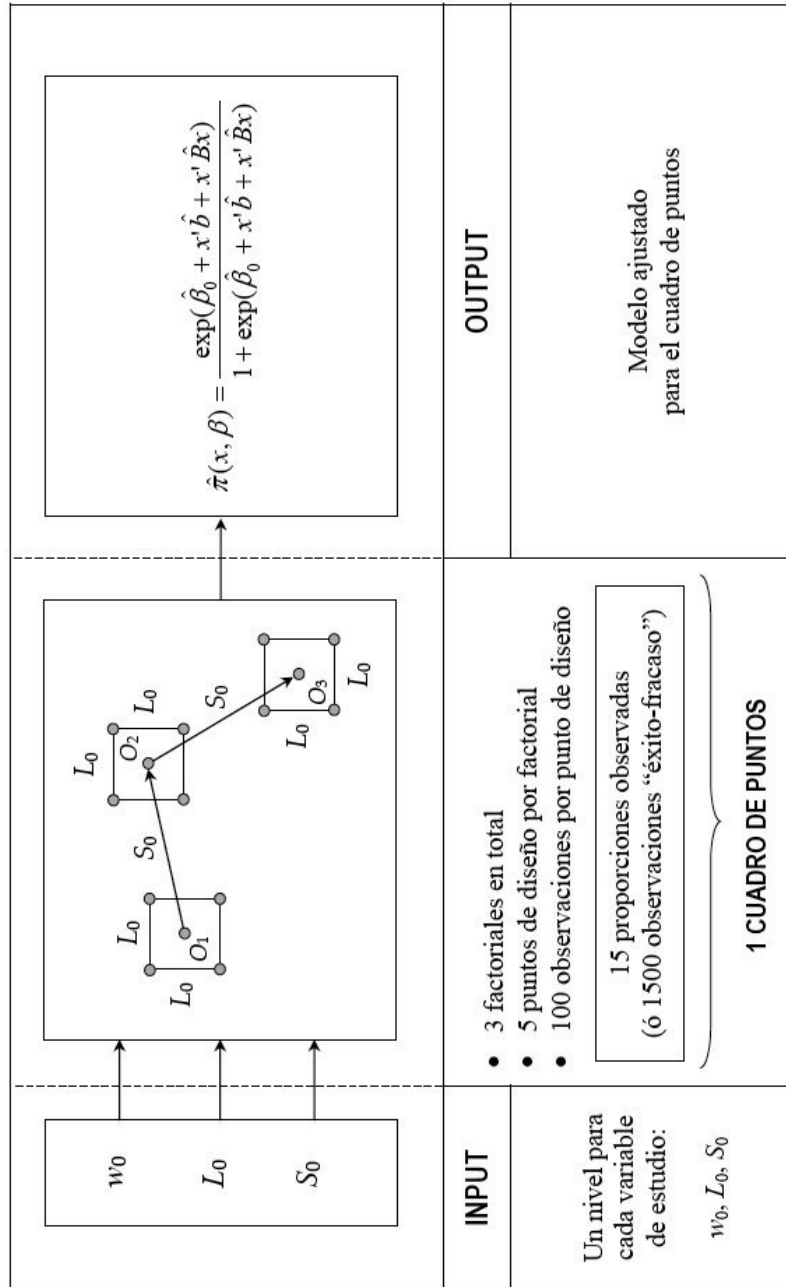


Figura 6.9: Esquema y definición del ajuste automático de un modelo para un *cuadro de puntos* compuesto por 15 puntos de diseño, generados aleatoriamente, y para un valor fijo de la terna de variables de estudio w_0 , L_0 y S_0 .