

SOFTWARE, INSTRUMENTACIÓN Y METODOLOGÍA

CURVAS NORMALES SUPERPUESTAS: ANÁLISIS GRÁFICO DEL ESPACIO PARAMÉTRICO

Concepción San Luis y Alfonso Sánchez Bruno

Universidad de La Laguna

Se propone el estudio de la representación gráfica del espacio paramétrico de la función de mínimos cuadrados de dos curvas normales superpuestas. En concreto se estudia el problema de la estimación de las alturas, de las medias y de las varianzas, suponiendo en cada caso conocidos los restantes parámetros. Los ejes de X e Y corresponden, según el parámetro estudiado en cada caso, a las alturas, medias y varianzas de la primera y segunda curva. Por su parte, el eje Z, simbolizado mediante curvas de nivel, corresponde a los valores que asume la función de suma de cuadrados para cada combinación de valores de los representados mediante los ejes X e Y. La utilidad de esta metodología es comprender los fenómenos que se dan en la estimación de los parámetros y, posiblemente, de descubrir características nuevas en los mismos.

Overlapping Normal Curves: Graphical Analysis of Parametric Space. It is proposed to study graphic representation of the parameter space of the minimum square function of two overlapping normal curves. In fact, it is studied the problem of estimating the heights, the means and the variances, assuming known the rest of the parameters in every case. The X and Y axes represent, relating to the parameter studied in every case, the heights, means and variances of the first and second curve. On the other hand, the Z axe, symbolised by contour levels, represents the values that assumes the sum of squares function for each combination of the values represented with the X and Y axes. The utility of this methodology is to understand the phenomena produced in the estimation of the parameters and, possibly, discovering new characteristics in them.

En el presente trabajo se propone el estudio de la representación gráfica del espacio

paramétrico de la función de mínimos cuadrados obtenida a partir de la estimación simultánea de dos de los parámetros de la función obtenida al superponer dos curvas normales diferentes. Esta propuesta se plantea como una herramienta destinada a la comprensión del comportamiento de los algoritmos de determinación de dichos parámetros.

Correspondencia: Concepción San Luis
Facultad de Psicología
Universidad de La Laguna
Tenerife (Spain)
E-mail: csanluis@ull.es

Los estudios basados en técnicas de análisis gráficos cobran cada día mayor importancia, siendo muchos los autores que los recomiendan efusivamente (ver, por ejemplo, Hartwig y Dearing, 1984).

Una de las razones que han llevado a su popularización es el desarrollo de las posibilidades gráficas de la microinformática, así como su cada día mayor potencia de cálculo, que facilitan el análisis y representación gráfica de funciones extremadamente complejas.

Se entiende por «espacio paramétrico» la representación de los valores de una función en relación a los de sus parámetros bajo el supuesto de que los valores de las variables (dependientes e independientes) son fijos (Draper y Smith, 1981). En el caso de que la función en cuestión posea sólo dos parámetros se obtendrá una representación tridimensional en la que sobre los ejes X e Y se representarán una serie de valores de cada uno de los parámetros y sobre el eje Z los valores de la función.

En el caso que nos ocupa, dos curvas normales superpuestas, se ha estudiado el problema de la estimación de las alturas, de las medias y de las varianzas, suponiendo en cada caso conocidos los restantes parámetros. Así, los ejes X e Y corresponden, según el gráfico, respectivamente a las alturas, medias y varianzas de la primera y segunda curva. Por su parte, el eje Z, simbolizado mediante curvas de nivel, corresponde a los valores que asume la función de suma de cuadrados para cada combinación de valores de los representados mediante los ejes X e Y.

Método

Para la realización de las figuras se generó un vector \vec{v} de 1.000 valores, comprendidos entre -3 y 3, igualmente distanciados entre sí. A partir de este vector se generó otro $w\vec{\omega}$, cuyos valores se calcularon según:

$$\omega_i = H_1 \cdot \exp\left(-\frac{1}{2} \frac{(v_i - \mu_1)^2}{\sigma_1^2}\right) + H_2 \cdot \exp\left(-\frac{1}{2} \frac{(v_i - \mu_2)^2}{\sigma_2^2}\right)$$

donde H_1 , μ_1 y σ_1^2 corresponden respectivamente a la altura en la media ($H_1=1$), la media ($\mu_1=-1$) y la varianza ($\sigma_1^2=1$) de la primera curva, y H_2 , μ_2 y σ_2^2 son sus homólogos en la segunda curva con valores 1, 2 y 1 respectivamente.

Para la realización de la primera figura se supusieron conocidas las medias y las varianzas y desconocidas las alturas. Así pues, se generaron dos vectores, \vec{H}_{1e} y \vec{H}_{2e} , de 250 valores cada uno, equidistribuidos entre -3 y 3, correspondientes a valores «supuestos» para las alturas de ambas curvas, pues, aún cuando una altura no puede, evidentemente, ser negativa, la finalidad de la investigación es estudiar el espacio paramétrico de una función y no sólo las soluciones que resulten lógicas desde una interpretación determinada de la función.

Para cada pareja de valores se generó un vector $\vec{\omega}_e$ de forma similar a $\vec{\omega}$, si bien sustituyendo H_1 y H_2 por la correspondiente pareja de valores de \vec{H}_{1e} y \vec{H}_{2e} , y se calculó la función de suma de cuadrados: $S = \sum\{\omega_i - \omega_{ie}\}^2$.

De esta forma, se obtuvo para cada figura una matriz de 250 x 250 valores de suma de cuadrados que se pueden representar, mediante curvas de nivel, como la «altura» correspondiente a cada uno de los puntos de un plano, cuyos ejes horizontales corresponden a \vec{H}_{1e} y \vec{H}_{2e} .

Evidentemente, el punto más bajo de dicha representación corresponderá, o estará cercano a, los verdaderos valores de H_1 y H_2 , si bien, en principio, podrían también aparecer máximos, mínimos locales y puntos de inflexión.

Las figuras 2 y 3 se realizaron de idéntica forma, si bien en la figura 2 se supusieron desconocidas las medias y conocidas las alturas y varianzas y en la figura

3 se supusieron desconocidas las varianzas y conocidas las medias y las alturas. Además, como se comentará posteriormente, en la figura 3 el rango de valores de las varianzas no iba de -3 a 3, sino de 0 a 3.

La programación se realizó en Gauss, si bien se modificó el multiplicador del generador de números aleatorios, tomando ahora 1343714438, que en estudios anteriores (Sánchez Bruno y San Luis, 1995) demostró tener un óptimo comportamiento.

Resultados

Aún cuando se ha denominado a esta sección «resultados», sería posiblemente mejor denominarla «ejemplos» pues, en sentido estricto, constituyen únicamente ejemplos de la aplicación del método de análisis gráfico que se propone.

En la figura 1 se ha representado el primer gráfico, en el cual, como se ha dicho anteriormente, se suponen desconocidas las alturas de ambas curvas. Como se puede

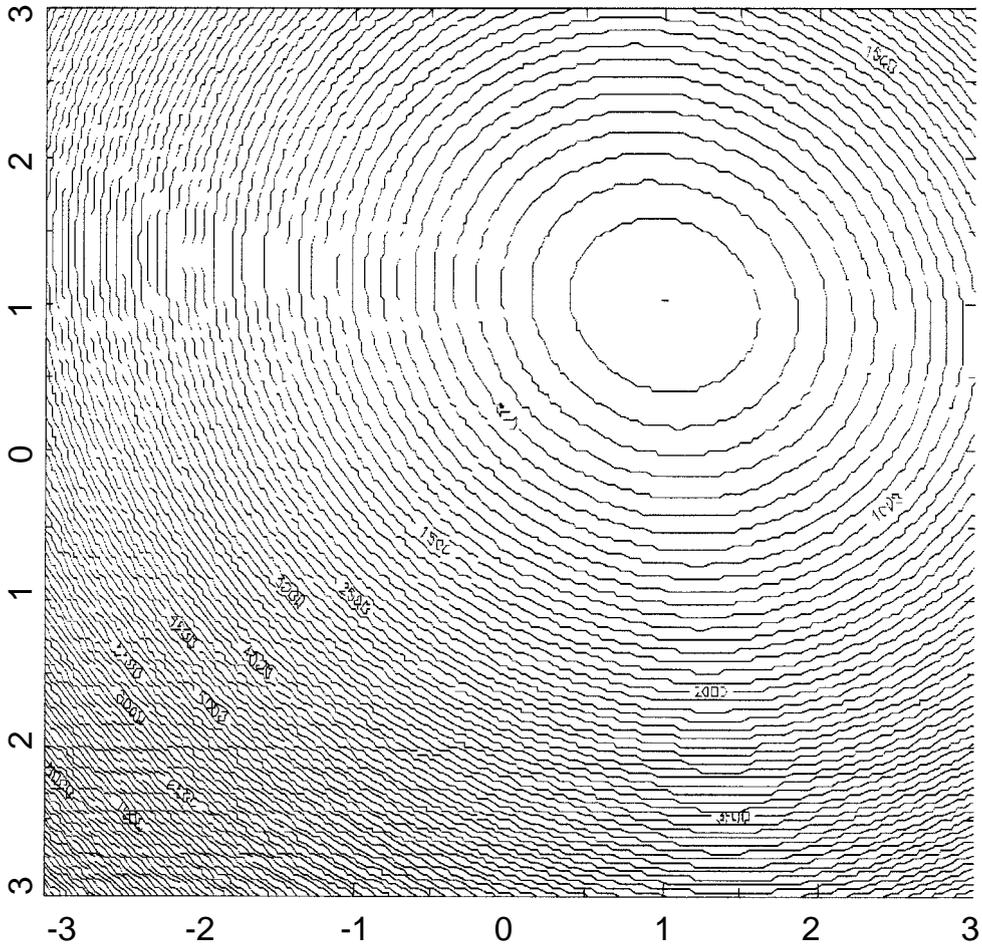


Figura 1. Alturas supuestamente desconocidas. Restantes parámetros conocidos.

ver, existe un punto mínimo y muy cercano al punto de valores reales, al que tienden a converger todas las líneas de pendiente, no siendo evidentes en el gráfico ni mínimos locales, ni máximos ni puntos de inflexión. Por consiguiente, la determinación de las alturas siendo conocidos los restantes parámetros no presenta ninguna dificultad, pues cualquier algoritmo de estimación convergerá en dicho mínimo.

En la figura 2 se han supuesto desconocidas las medidas y conocidos los demás pa-

rámetros. Como se puede apreciar, junto al verdadero «mínimo» (situado en el cuadrante superior izquierdo de la figura y correspondiente a los valores verdaderos de los parámetros), existen otro mínimo (imagen especular del «real»), dos máximos (en los cuadrantes inferior izquierdo y superior derecho), y un punto de inflexión centrado con respecto a los cuatro puntos críticos antes descritos.

Dada la complejidad del espacio paramétrico, parece evidente que la convergencia a

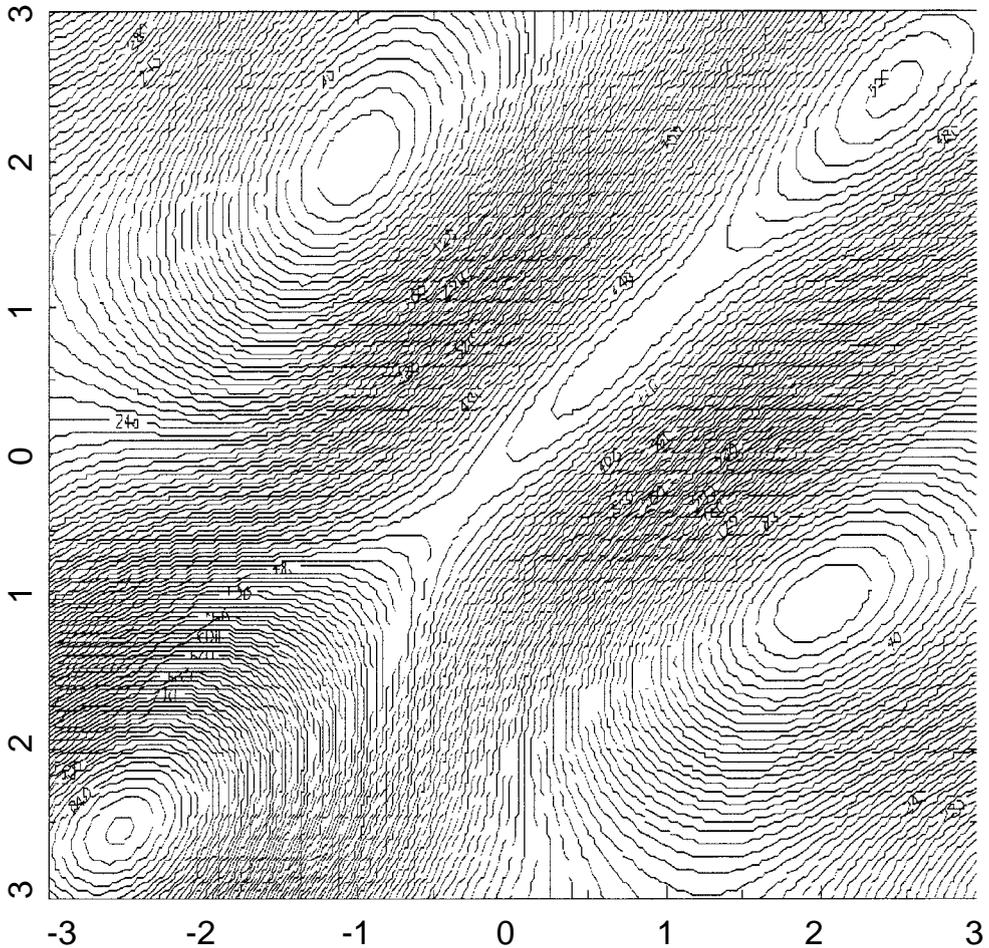


Figura 2. Medias supuestamente desconocidas. Restantes parámetros conocidos.

los verdaderos valores de las medias será imposible a menos que el algoritmo de estimación (necesariamente iterativo dada la naturaleza no lineal del problema) parta de un punto muy cercano al final.

Démonos cuenta, además, de que la situación que aquí se ha presentado está enormemente simplificada, pues además de las medias (por otra parte en número casi siempre superior a dos), suele necesitarse estimar, como poco, las alturas, con lo que el espacio paramétrico se complejiza aún más.

Por último, la figura 3 es diferente de las anteriores, pues los valores de los ejes son sólo positivos (rango de valores entre 0 y 3). Aún cuando, como se había dicho antes, se admitirían valores absurdos para los parámetros, el aceptar varianzas negativas supone unos valores tan enormes de la función suma de cuadrados que, siendo imposible la convergencia a ningún punto en dicha zona por carecer ésta de puntos críticos, se dificulta la apreciación del resto de la figura.

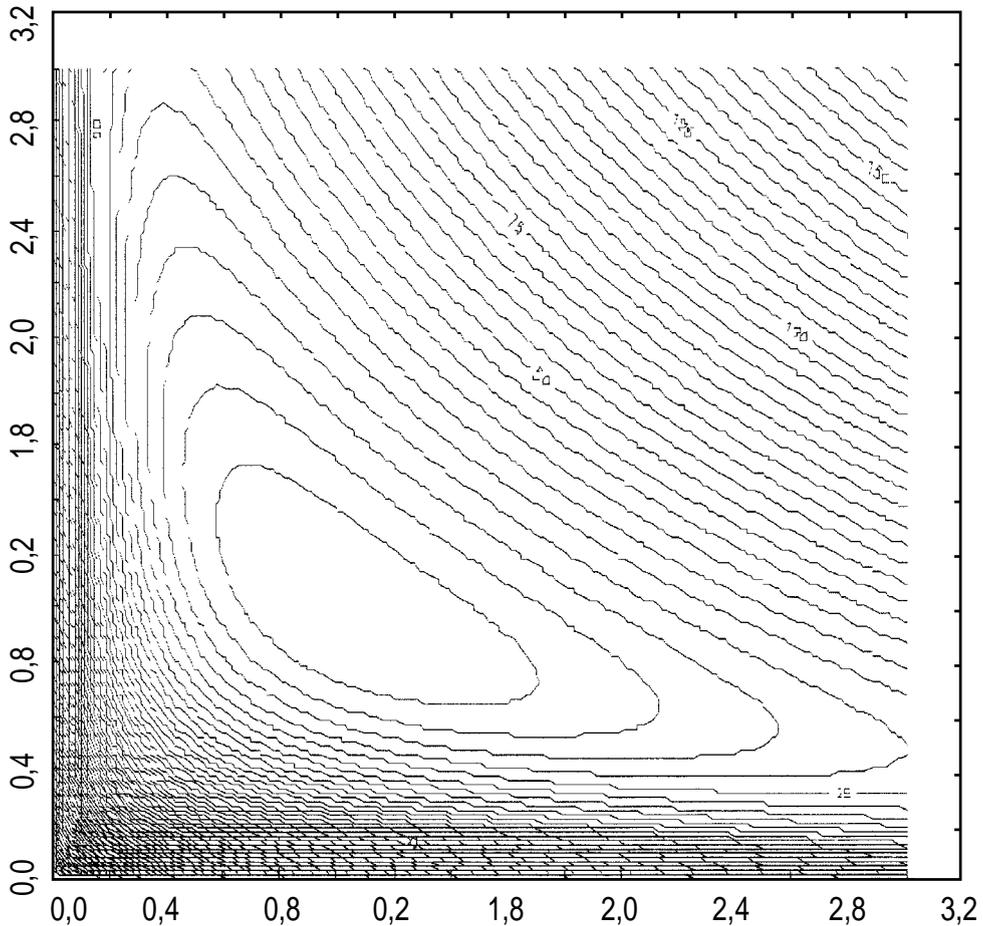


Figura 3. Varianzas supuestamente desconocidas. Restantes parámetros conocidos.

Tal como se observa, aún cuando esta estimación es eminentemente no lineal, no debe ser difícil en principio la convergencia a los verdaderos valores de las varianzas.

Finalmente, se debe insistir en que la propuesta que aquí se hace no está orienta-

da tanto al descubrimiento como a la comprensión. Desgraciadamente, el método se complica enormemente cuando se trabaja con más de dos parámetros, pero este problema se podría solucionar en parte mediante proyecciones o mediante cortes transversales al nuevo eje.

Referencias

Draper, N. y Smith, H. (1981). Applied regression analysis, second edition. U.S.A.: Wiley.
Hartwig, F. y Dearing, B.E. (1984). Exploratory Dta Analysis. Inglaterra: Sage Publications.
Sánchez Bruno, A. y San Luis, C. (1994) «A statistical Analysis of Seven Multipliers for

Linear Congruential Random Number Generators with Modulus - 1». *Quality & Quantity*, 29(3), 331-337.

Aceptado el 22 de abril de 1998