gelands under large-scale commercial grazing. *Ecological Applications*, **4** (3): 497-517.

Prince, S.D. (2002): Spatial and temporal scales for Detection of Desertification. In J.F. Reynolds and D.M. Stafford Smith (eds.). Global Desertification, Do Humans cause Deserts? Dahlem Workshop Report 88. Dahlem University Press.

Puigdefabregas, J. & Mendizábal, T. (2004): Prospects of desertification impacts in Western Europe. In: A. Marquina (ed.). Environmental Challenges in the Mediterranean 2000.2050. NATO Science Series. IV. Earth and Environmental Sciences, **37**: 155-174.

Puigdefabregas, J. & Mendizábal, T. (2006): Desertification and migrations in Western Mediterranean: Environmental outcomes of the inverse flow of people and capital. 2nd International Symposium on desertification and migrations, United Nations Convention to Combat Desertification (UNCCD) Proceedings (full texts in CD). Ministerio de Medio Ambiente, Dirección General de la Biodiversidad, Madrid.

Puigdefabregas, J. (1995): Desertification: Stress beyond resilience, exploring a unifying process structure. *Ambio*, **24**: 311–313.

Puigdefabregas, J. (1998): Variabilidad climática y sus consecuencias sobre la sostenibilidad de los sistemas agrarios. In: *Agricultura Sostenible*. Jiménez Díaz, R. & Lamo de Espinosa, J. (eds). 41-70. Ediciones Mundi-Prensa. Madrid.

Puigdefabregas, J. (2005): Desertificación, un fenómeno global que afecta a España. *Acta Científica y Tecnológica*, **9**: 19-23. Madrid.

Thebaud, B. (1993): Causes et consequences de la désertification au Sahel de l'Ouest: Essai d'interpretation. International Panel of Experts. Secretariat of the Intergovernmental Negotiating Committee for a Convention to Combat Desertification. INCD-UN, p. 7. Geneva.

UNCCD. United Nations Convention to Combat Desertification in Countries Experiencing Serious Drought and/or Desertification, Particularly in África. UNCCD Secretariat. Bonn (available at http://www.unccd.int).

Wiegand, T., Milton, S. & Wissel, C. (1995): A simulation model for shrub ecosystem on the semi-arid karoo, South Africa. Ecology, **7**: 2205-2022.

Juan Puigdefabregas
Estación Experimental de Zonas Áridas (CSIC)

Matemáticas

Estadística Bayesiana: la incorporación de la experiencia al análisis de datos

En este trabajo se pretende describir los conceptos básicos de la *estadística bayesiana*. Su nombre hace referencia al conocido teorema de Bayes, en el que esta metodología tiene su fundamento y su origen. Se trata de una sencilla fórmula matemática que muestra, en términos de probabilidades, cómo se conjuga la observación de unos datos con la información que, por la experiencia previa, se tenga sobre las posibles causas que los producen.

El teorema de Bayes se debe al matemático inglés Thomas Bayes y fue publicado en 1763.

Thomas Bayes (1702-1761)

A partir de ese germen inicial, se fue desarrollando la metodología bayesiana, hasta llegar a constituir hoy una rama de la matemática con entidad propia. Se celebran periódicamente congresos internacionales de metodología bayesiana, en los que intercambian resultados muchos de los especialistas que trabajan en ella. Son especialmente significados los *Valencia International Meetings on Bayesian Statistics*, organizados por el Profesor J. M. Bernardo y copatrocinados por la Universidad de Valencia y la *International Society for Bayesian Analysis* (ISBA).

Bayesian Statistics
VALENCIA 8
ISBA 2006

Logotipo de Valencia 8

1. EL PLANTEAMIENTO BAYESIANO

En un problema de inferencia estadística hay una cantidad, un *parámetro*, desconocido por el investigador, del que se sabe que pertenece a un determinado conjunto, llamado *espacio paramétrico*, y se pretende estimar su valor o incluso tratar de responder algunas preguntas formuladas sobre el mismo. Para obtener alguna información sobre este parámetro desconocido, se observa un conjunto de datos. Los datos son observaciones de una *muestra aleatoria*, o sea, un vector aleatorio, cuya distribución de probabilidad está relacionada con el valor del parámetro. La función de masa o de densidad de la muestra depende del parámetro; a esta función, una vez que la muestra ha sido observada, se la considera función del parámetro, y se la llama función de verosimilitud. El objeto matemático que engloba y precisa matemáticamente la información que la muestra suministra sobre el parámetro es esta *función de verosimilitud*.

La estadística bayesiana se caracteriza por añadir a este planteamiento básico los dos supuestos siguientes:

- Antes de observar la muestra se tiene alguna información sobre el parámetro.
- Esta información se expresa mediante la asignación de una distribución de probabilidad, llamada *distribución inicial* o *a priori*.

En efecto, se supone que antes de observar la muestra se tiene ya, debido a la experiencia anterior, alguna información previa acerca del parámetro, que también debe ser aprovechada para llegar a un mejor conocimiento del mismo. Para designar un objeto matemático que represente esta información, se considera al parámetro como una *variable aleatoria subjetiva*, y se le asigna una distribución de probabilidad, determinada por el investigador, o su consultante, en la que se reflejen los conocimientos previos de éste; esta distribución de probabilidad es el objeto matemático que representa la información previa: es la *distribución inicial* o *a priori*.

Las dos informaciones, la previa a la observación y la suministrada por la muestra, se acumulan dando lugar a una información total. Esta información total es representada por otro objeto matemático: la llamada *distribución final* o *a posteriori* del parámetro, que se define como la distribución del parámetro condicionada por la muestra.

Veamos con más detalle la matemática de este proceso, introduciendo, al mismo tiempo, la notación usual. Por no duplicar expresiones similares, consideraremos que tanto el parámetro como la muestra son variables continuas; en el caso discreto, bastaría cambiar la expresión «función de densidad» por «función de masa» y las integrales por sumatorios.

(a) Existe un parámetro θ, que pertenece a un espacio paramétrico Θ, que, a su vez, es un subconjunto de $\mathbb{R}^k$, para algún k. Esto es: $\theta \in \Theta \subset \mathbb{R}^k$. Como vemos, el parámetro puede ser unidimensional, por tanto, un número (si $k = 1$), o multidimensional, es decir, un conjunto de números o un vector (si $k \geq 2$).

(b) Para el parámetro θ se tiene una función de *densidad inicial*, subjetiva, $\pi(\theta)$, definida en el espacio paramétrico Θ. (Esta densidad representa la información previa a la muestra.)

(c) Se considera una muestra, que es una variable aleatoria (un vector aleatorio) $(X_1,\ldots, X_n)$ de dimensión o *tamaño n*.

(d) Para cada valor θ del parámetro, dentro de Θ, la función de densidad, condicionada por θ, de cada una de las n componentes $X_1,\ldots,X_n$ de la muestra es $f(\cdot \mid \theta)$, la misma para todas ellas. A esta densidad se la llama a veces *densidad poblacional*. Las componentes X_i de la muestra, una vez condicionadas por θ, son independientes entre ellas.

(e) La densidad de toda la muestra $X_1,\ldots,X_n$ condicionada por θ, denotada también usualmente por f, es:

$$f(x_1,\ldots,x_n \mid \theta) = f(x_1 \mid \theta)\ldots f(x_n \mid \theta)$$

Esta función (considerada como función de θ) es la *función de verosimilitud*. (Esta densidad representa la información muestral).

(f) La *densidad final* o *a posteriori* del parámetro es su densidad condicionada por la muestra particular $(x_1,\ldots,x_n)$, esto es,

$$\pi(\theta \mid x_1,\ldots,x_n) = \frac{\pi(\theta) f(x_1,\ldots,x_n \mid \theta)}{\int_\Theta \pi(\theta) f(x_1,\ldots,x_n \mid \theta)\, d\theta} \quad (1)$$

(Esta densidad representa la información total).

La expresión (1) es precisamente el teorema de Bayes, conocido ya por nuestros alumnos de bachillerato, sólo que aquí está expresado en términos de distribuciones continuas.

Dado que el numerador del último término de (1) no depende de θ, podemos escribir simplemente:

$$\pi(\theta \mid x_1,\ldots,x_n) \propto \pi(\theta) f(x_1,\ldots,x_n \mid \theta), \quad (2)$$

donde el símbolo $\propto$ indica proporcionalidad; expresamos así la densidad final prescindiendo de una constante normalizadora, que la explicitaría totalmente.

La fórmula (2) permite identificar la influencia de las dos informaciones de entrada en la información final. En efecto, una persona que sólo tuviera la información previa, anterior a la muestra, se inclinaría hacia cada posible valor θ con una fuerza proporcional a su densidad o probabilidad inicial $\pi(\theta)$; un estadístico no bayesiano, que sólo considerara la información muestral, se inclinaría hacia cada valor θ con una fuerza proporcional a su verosimilitud $f(x_1,\ldots,x_n \mid \theta)$; un estadístico bayesiano, que considera ambas informaciones, se inclina a θ con una fuerza que es proporcional simultáneamente a las dos anteriores y, por tanto, proporcional al producto de ambas.

Hemos de añadir que en la actualidad se considera la existencia de distribuciones iniciales $\pi(\theta)$ «objetivas», que reducen la influencia del observador y hacen que el resultado dependa básicamente de los datos.

Un ejemplo

Apliquemos estas consideraciones generales a un ejemplo.

Supongamos que estamos interesados en el conocimiento de la magnitud *«proporción de hogares de la ciudad de Bayesville que tienen conexión a Internet»*. Nuestro parámetro θ es esa proporción desconocida; el espacio paramétrico al que pertenece esta proporción es el intervalo (0, 1); esto es: $\theta \in \Theta = (0, 1)$. Para obtener información sobre θ vamos a elegir al azar un número n de hogares de Bayesville y preguntaremos si tienen conexión a Internet. Ésa es una de nuestras dos fuentes de información: la muestra.

Antes de hacer la encuesta, para nosotros la respuesta de cada uno de los hogares es desconocida, depende del azar. Por ello, a cada uno de los n hogares le tenemos asignada una variable aleatoria (unidimensional) a la que le haremos tomar el valor 1 si la respuesta del hogar es SÍ y el valor 0 si la respuesta es NO. El conjunto de estas n variables $(X_1,...,X_n)$, donde cada X_i, = 0 ó 1, es una muestra aleatoria. Si la proporción de hogares con conexión es θ, la probabilidad de que las variables binarias X_i tomen el valor 1 es θ, y para el valor 0 es $1 - \theta$. Por lo tanto, su función de masa es, para todas ellas,

$$f(1 \mid \theta) = \theta; \quad f(0 \mid \theta) = 1 - \theta$$

Ésta es la función de masa poblacional, que podemos escribir también así:

$$f(x \mid \theta) = \theta^x (1-\theta)^{1-x}, \quad \text{para } x \in \{0, 1\}$$

La información suministrada por nuestra muestra particular de ceros y unos se contiene matemáticamente en la función de verosimilitud, que es la función de masa de la muestra entera condicionada por θ. Se calcula fácilmente que esta función es:

$$f(x_1,\ldots,x_n \mid \theta) = \theta^{\Sigma x_i} (1-0)^{n-\Sigma x_i} \tag{3}$$

Esto en cuanto a la información suministrada por la muestra. Pero hemos visto que en la estadística bayesiana se considera también la información que la experiencia previa ofrece acerca de θ. Esta información depende de lo que conozca el investigador, o su consultante, y se plasma en la distribución inicial. En este ejemplo es razonable pensar que una distribución Beta (μ, τ) de la familia Beta, que abarca una amplia gama de distribuciones, todas sobre el intervalo (0, 1), puede reflejar bien los conocimientos del consultante; así, sólo hay que determinar el valor que se desea que tengan las constantes μ y τ de esta distribución. (Las constantes μ y τ que hemos elegido son diferentes, aunque lógicamente equivalentes, a las que se usan habitualmente para identificar la densidad Beta; la ventaja es que estas constantes tienen un claro significado probabilístico, como veremos enseguida). Supongamos que después de analizar el caso (más adelante nos referiremos a ello) se ha elegido una distribución Beta (μ, τ), con unos valores μ y τ determinados. Con ello, la correspondiente función de densidad inicial, expresada sólo en términos de proporcionalidad, es:

$$\pi(\theta) \propto \theta^{\mu\tau-1}(1-\theta)^{(1-\mu)\tau-1}, \quad \text{para } \theta \in (0, 1) \tag{4}$$

De las propiedades de la Beta resulta que la esperanza y varianza de la distribución inicial o a priori de θ son:

$$\mathrm{E}[\theta] = \mu; \quad \mathrm{V}[\theta] = \frac{\mu(1-\mu)}{\tau+1} \tag{5}$$

La constante μ es, por tanto, una característica de centralización y la constante τ puede interpretarse como una medida de precisión (inverso de la varianza).

Conviene señalar que las fórmulas en (5) son las que nos dan la idea de cómo se puede escoger entre las distribuciones Beta la distribución inicial o a priori que represente las opiniones del consultante: para ello preguntamos a éste cuál considera él que es el verdadero valor de la proporción θ y asignamos a μ ese valor; después le preguntamos qué confianza tiene en la conjetura que acaba de proponer para θ y así asignamos a τ un valor tanto mayor cuanto mayor sea esa confianza. El valor de μ es, pues, la opinión sobre θ y el valor de τ mide la calidad o fuerza informativa de esa opinión.

Pongamos números concretos. Supongamos que al preguntar al consultante cuál conjetura él que puede ser el valor de θ nos dice que 0.4: hacemos $\mu = 0.4$. Supongamos que a base de preguntarle qué grado de confianza tiene él en esa conjetura hemos visto que conviene hacer $\tau = 10$. En este caso tomamos como distribución inicial la Beta (0.4, 10). Su gráfica es la curva de trazo fino de la figura 1.

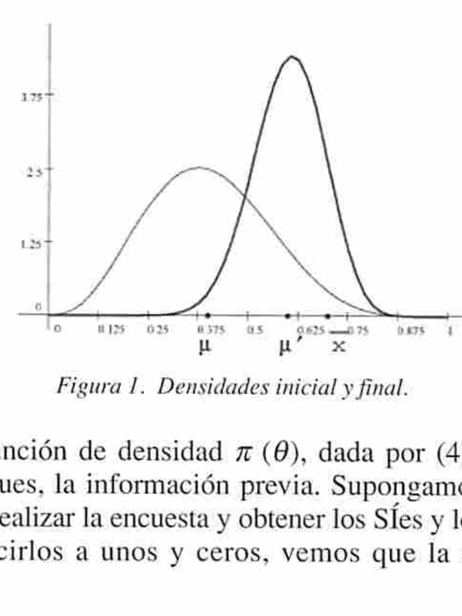

Figura 1. Densidades inicial y final.

La función de densidad $\pi(\theta)$, dada por (4), representa, pues, la información previa. Supongamos ahora que, al realizar la encuesta y obtener los SÍes y los NOes y traducirlos a unos y ceros, vemos que la muestra

$(X_1,...,X_n)$ ha tomado el valor $(x_1,...,x_n)$, donde cada x_i es 1 ó 0 según fuera la respuesta obtenida. De la muestra obtenemos la media muestral $\bar{x} = (\Sigma x_i)/n$, que es la proporción de hogares en la muestra obtenida que tienen conexión a Internet. Es claro que la muestra apunta al número $\bar{x}$ como el posible valor θ. Por otra parte, la calidad o fuerza informativa de $\bar{x}$ viene medida por el tamaño n de la muestra. Poniendo también números concretos, supongamos que se ha tomado una muestra de $n = 20$ hogares, y que de ellos 14 resultaron tener conexión a Internet y 6 no. La media o proporción muestral es $\bar{x}$ = 0.7. La muestra apunta, pues, al número 0.7 y su fuerza informativa es 20.

La densidad final, de acuerdo con (2), se obtiene multiplicando (4) por (3). Haciéndolo así, se obtiene que esta densidad final es también una Beta, pero con otras constantes diferentes, es Beta (μ', τ') con:

$$\mu' = \frac{\tau}{\tau + n}\mu + \frac{n}{\tau + n}\bar{x}; \qquad (6)$$

$$\tau' = \tau + n \qquad (7)$$

Si antes de observar los n hogares las opiniones del usuario sobre la proporción θ eran las reflejadas en la distribución inicial Beta (μ,τ), una vez observados los hogares, sus opiniones serán las reflejadas en la distribución final Beta(μ', τ').

El valor que conjeturamos ahora para θ, después de observar la muestra, ya no es μ, sino μ'. La fórmula (6) nos indica que ésta es una media ponderada entre la conjetura inicial μ y la proporción muestral $\bar{x}$, siendo las respectivas ponderaciones proporcionales a τ y n, esto es, a las correspondientes confianzas o calidades informativas.

La fuerza o calidad informativa de la nueva μ' como conjetura de θ ya no es τ sino τ'. La fórmula (7) nos dice que esta confianza es la suma de la confianza τ en la conjetura inicial más la confianza en la muestra n.

En nuestro ejemplo numérico es $\mu = 0.4$, $\bar{x} = 0.7$; $\tau = 10$ y $n = 20$; de ahí obtenemos que $\mu' = 0.6$ y $\tau' = 30$. La gráfica de la densidad a posteriori correspondiente, Beta (0.6, 30), es la curva de trazo grueso de la figura 1. Está a la derecha de la densidad inicial y es más estrecha y alta.

A partir de ahora, nuestra conjetura para θ ya no es $\mu = 0.4$, sino $\mu' = 0.6$; y su fuerza informativa ya no es $\tau = 10$, sino $\tau' = 30$.

En la figura 1 vemos que $\mu' = 0.6$ está al doble de distancia de $\mu = 0.4$ que de $\bar{x} = 0.7$. Ello se debe a que la fuerza informativa de $\bar{x}$ es $n = 20$, el doble de la de μ, que es $\tau = 10$. También hemos notado que la gráfica de la densidad final es más alta y estrecha que la de la inicial. Ello se debe a que la nueva fuerza informativa, τ' = 30, es mayor que la anterior, $\tau = 10$.

En nuestro ejemplo el tamaño de la muestra es $n = 20$. Si el tamaño fuera mucho mayor, ¿cómo sería la distribución final?

En este caso la fórmula (6) nos dice que la conjetura final μ' sería muy próxima a $\bar{x}$; la influencia de μ sería muy pequeña: la influencia de la distribución inicial va siendo menor.

La fórmula (7) nos dice que la confianza en μ' como nueva conjetura sería muy grande.

Por otra parte, el valor de $\bar{x}$ sería próximo al verdadero valor de θ. Luego la gráfica de la función de densidad estaría centrada cerca de θ y sería muy alta y estrecha. Nuestra conjetura μ' sería muy buena y nuestra confianza en ella sería muy grande.

2. INFERENCIA BAYESIANA

Tres problemas básicos en inferencia son la estimación por punto, la estimación por intervalo y los contrastes de hipótesis. Vamos a ver el modo bayesiano de abordar estos problemas.

2.1. Estimación por punto

En un problema de estimación por punto se trata de responder de algún modo a la pregunta ¿cuál es el valor del parámetro? La respuesta, desde el punto de vista bayesiano, es una estimación (esto es, un número), junto con una medida de la fiabilidad de esta estimación.

Un buen modo de hacer consiste en dar como estimación la esperanza final o a posteriori del parámetro, $E[\theta \mid x_1,...,x_n]$, y dar como medida de la fiabilidad, o más bien del error, la desviación típica final de θ, $D[\theta \mid x_1,...,x_n]$, que representa la distancia media de θ a su esperanza.

En nuestro ejemplo se obtiene que

$$E[\theta \mid x_1,\ldots,x_n] = \mu' = 0.6;$$

$$D[\theta \mid x_1,\ldots,x_n] = \sqrt{\frac{\mu'(1-\mu')}{\tau}} = 0.088$$

Con la distribución inicial habríamos obtenido la estimación 0.4 y la medida del error 0.148.

2.2. Estimación por intervalo

La estimación por intervalo, desde el punto de vista bayesiano, consiste sencillamente en mostrar un intervalo, llamado *intervalo creíble*, tal que la probabilidad a posteriori de que el parámetro caiga en él tome un valor alto, $1 - \alpha$, previamente fijado. Esto se consigue con cierta facilidad a partir de la distribución a posteriori.

En general se desea, además, que el intervalo creíble tenga la menor longitud posible. Para ello se toma, entre todos los intervalos que tengan la probabilidad fijada, el que incluye los puntos en los que la densidad a posteriori es más alta. Es el llamado intervalo HPD (high posterior density).

En nuestro ejemplo, si deseamos un intervalo creíble de probabilidad $1 - \alpha = 0.95$, obtenemos que el intervalo HPD correspondiente es (0.427, 0.769). La probabilidad a posteriori de que la proporción de hogares con conexión esté entre 0.427 y 0.769 es de 0.95. Es el área entre las dos líneas verticales de la figura 2.

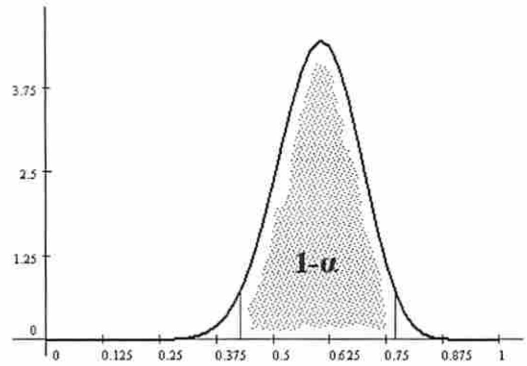

Figura 2. Intervalo creíble HPD.

2.3. Contrastes de hipótesis

En un problema de contraste de hipótesis se plantea la pregunta de si el parámetro θ estará dentro de un conjunto fijado por el investigador o su consultante o si, por el contrario, estará fuera de ese conjunto. La respuesta consiste sencillamente en calcular la probabilidad final o a posteriori de cada una de las dos hipótesis.

En nuestro ejemplo, supongamos que el consultante considera que sería bueno abrir una nueva tienda de informática en Bayesville si la proporción θ de hogares con conexión fuera mayor de 0.5. La pregunta es ¿estará el parámetro en el intervalo (0.5, 1) o estará fuera de él? La respuesta consiste en dar las probabilidades a posteriori correspondientes, que resultan ser:

$$\alpha_0 = \mathrm{P}(\theta \in (0.5, 1) \mid x_1, \ldots, x_n) = 0.8675$$
$$\alpha_1 = \mathrm{P}(\theta \in (0, 0.5) \mid x_1, \ldots, x_n) = 0.1325$$

Son las áreas de las zonas a la derecha y a la izquierda, respectivamente, de la línea vertical en la figura 3.

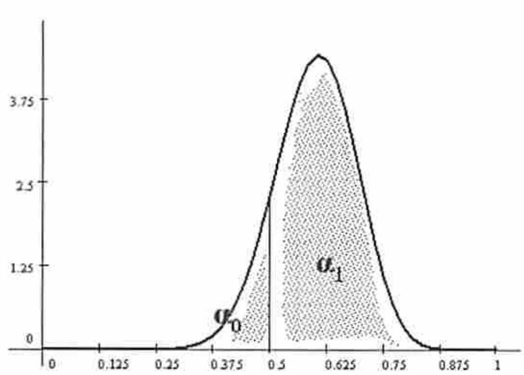

Figura 3. Probabilidades finales de la hipótesis.

Con esta información y sus conocimientos del mercado y de su situación empresarial, el consultante puede estimar si le conviene o no abrir la nueva sucursal en Bayesville.

3. ASPECTOS COMPUTACIONALES: MÉTODOS MCMC

El aspecto computacional ha supuesto un freno, durante cierto periodo de tiempo, para la aplicación de la metodología bayesiana en situaciones prácticas. No obstante, los métodos *MCMC* (acrónimo de «Markov Chain Monte Carlo»), aparecidos recientemente, que permiten simular observaciones y estimar características de la distribución a posteriori sin tener que utilizar su expresión matemática, han impulsado enormemente el desarrollo de las aplicaciones de los métodos bayesianos.

Los dos métodos más utilizados son:

- el algoritmo de *Metrópolis-Hastings*;
- el *muestreador de Gibbs* (Gibbs sampler).

El algoritmo de *Metrópolis-Hastings* empieza generando observaciones de otra distribución (distribución de prueba), distinta de la que pretendemos analizar pero que se puede simular con cierta facilidad, y para cada observación simulada se realiza un sorteo mediante el cual se decide si se la considera observación de la distribución de interés o no, en cuyo caso se la rechaza.

El *muestreador de Gibbs*, por su parte, se aplica cuando el parámetro es multidimensional, y se basa en la sucesiva simulación de todas las distribuciones condicionadas univariantes. Así, si se trata de obtener observaciones de la distribución a posteriori, llamémosla $h\ (\theta_1, \ldots, \theta_k)$, el proceso consiste en establecer un valor inicial $\theta_1^0 \ldots, \theta_k^0$ y luego ir repitiendo etapas similares, en cada una de las cuales se siguen los siguientes pasos:

Paso 1: Obtener una observación de θ_1 según la distribución condicionada:

$$h(\theta_1 \mid \theta_2^0, \ldots, \theta_k^0) \rightarrow \theta_1^1$$

$$\vdots$$

Paso 2: Obtener una observación de θ_2 según la distribución condicionada:

$$h(\theta_2 \mid \theta_1^1, \theta_3^0, \ldots, \theta_k^0) \rightarrow \theta_2^1$$

Paso k: Obtener una observación de θ_k según la distribución condicionada:

$$h(\theta_k \mid \theta_1^1, \theta_2^1, \ldots, \theta_{k-1}^1) \rightarrow \theta_k^1$$

Las primeras observaciones del proceso de simulación han de ser desechadas, para anular el efecto del valor inicial $\theta_1^0 \ldots, \theta_k^0$, que es elegido arbitrariamente. A esta fase inicial se la denomina de «caletamiento» (burn-in) y su longitud suele determinarse por métodos gráficos que permiten visualizar la convergencia de las simulaciones. El resto de observaciones constituye la muestra efectiva.

A modo de ejemplo, si se desea simular observaciones de una densidad $h\ (\theta_1, \theta_2)$ normal bivariante:

$$N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

mediante el *muestreador de Gibbs*, se determina primero un valor inicial θ_1^0, θ_2^0, arbitrario y, a continuación, en cada etapa t se obtiene una observación (θ_1^t, θ_2^t), partiendo del valor $(\theta_1^{t-1}, \theta_2^{t-1})$ de la etapa $t-1$, mediante la simulación en los dos pasos siguientes:

$$\theta_1 \mid \theta_2 = \theta_2^{t-1} \quad \text{que es} \quad N_1(\rho\theta_2^{t-1}, 1-\rho^2) \rightarrow \theta_1^t$$
$$\theta_2 \mid \theta_1 = \theta_1^{t} \quad \text{que es} \quad N_1(\rho\theta_1^{t}, 1-\rho^2) \rightarrow \theta_2^t$$

Como ejemplo concreto, suponiendo $\rho = 0.80$, se ha aplicado el muestreador de Gibbs comenzando en $(\theta_1^0 \ldots, \theta_2^0) = (10{,}10)$ y se han obtenido las 50 observaciones representadas en el gráfico de la figura 4.

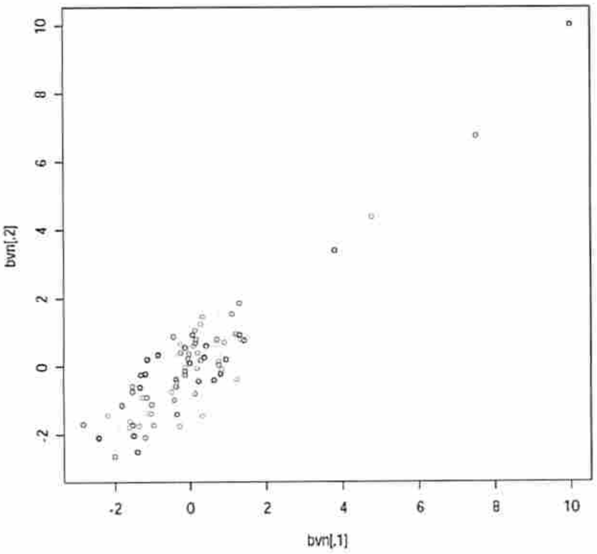

Figura 4. Muestreador de Gibbs.

Observando el gráfico se ve que conviene desechar las 5 primeras observaciones y quedarnos con una muestra efectiva de tamaño 45.

UNA APLICACIÓN A UN PROBLEMA DE DIETAS

Con la siguiente aplicación no pretendemos realizar un análisis bayesiano completo, sino sólo una ilustración escueta de los métodos computacionales.

En una granja, se aplican cuatro dietas diferentes a cuatro grupos de pollitos de la misma edad para comparar sus pesos. El modelo utilizado considera las observaciones Y_{ij}, $i = 1, \ldots, n$ para $j = 1, 2, 3, 4$, independientes y, para cada grupo, distribuidas normalmente con diferentes medias y con varianza común, es decir, $N\ (\theta_j, \sigma^2)$, $j = 1, 2, 3, 4$, por lo que las medias aritméticas $\overline{Y}_j$ serán $N\left(\theta_j, \dfrac{\sigma^2}{n_j}\right)$. Se toma como distribución a priori sobre cada θ_j una normal $N\ (230, 1)$; para σ^2 se toma una distribución a priori uniforme $U\ (0, 100)$.

Los pesos medios por dieta, obtenidos en 45 pollos, fueron:

	Dieta 1	*Dieta 2*	*Dieta 3*	*Dieta 4*
n_j	16	10	10	9
$\bar{y}_j$	177750	214700	270300	238556

A estos datos se les aplicaron métodos MCMC contenidos en el software libre de cálculo bayesiano WINBUGS. Así se simularon 10000 observaciones y se obtuvieron varias características de las distribuciones a posteriori de los parámetros θ_j. En particular, las medias a posteriori fueron éstas:

Parámetro	θ_1	θ_2	θ_3	θ_4
Medias a posteriori	224.9	225.0	225.1	225.0

No obstante, se puede obtener una descripción más completa del comportamiento de los parámetros utilizando toda la distribución a posteriori descrita mediante la simulación anterior; de este modo se pueden estimar medidas de dispersión, cuantiles e intervalos de probabilidad fija, que permitirán establecer de una forma ajustada las diferencias entre los efectos de las cuatro dietas.

Eusebio Gómez Sánchez-Manzano y
Paloma Maín Yaque
Dpto. de Estadística e Investigación Operativa
Facultad de Ciencias Matemáticas (UCM)