Modelo de error de medida como problema de decisión bayesiana no paramétrico

por VICENTE QUESADA PALOMA
y ALFONSO GARCIA PEREZ
Departamento de Estadística Matemática
Facultad de Ciencias
Universidad Autónoma de Madrid

1. PROCESOS DE DIRICHLET

Definición 1.1:

Sean $X_1, \ldots, X_n$ n variables aleatorias independientes, de forma que cada variable X_j sigue una distribución gamma $G(\alpha_j, 1)$, para $j = 1, \ldots, n$ y con $\alpha_j \geq 0$, existiendo algún i, de forma que $\alpha_i > 0$.

Consideremos las variables aleatorias

$$Y_j = \frac{X_j}{\sum_{k=1}^{n} X_k}, \quad j = 1, \ldots, n$$

A la distribución de la variable aleatoria n-dimensional $(Y_1, \ldots, Y_n)$ se le denomina distribución de Dirichlet de parámetros $(\alpha_1, \ldots, \alpha_n)$, la cual representaremos por $D(\alpha_1, \ldots, \alpha_n)$.

Si $\alpha_j > 0$ para $j = 1, \ldots, n$, la variable aleatoria $(n-1)$-dimensional $(Y_1, \ldots, Y_{n-1})$ es absolutamente continua con respecto a la medida de Lebesgue en $\mathbb{R}^{n-1}$ y tiene por función de densidad

$$f(y_1, \ldots, y_{n-1}/\alpha_1, \ldots, \alpha_n) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_n)}{\Gamma(\alpha_1) \cdot \ldots \cdot \Gamma(\alpha_n)} \cdot \left(\prod_{i=1}^{n-1} y_i^{\alpha_i - 1} \right) \cdot \left(1 - \sum_{i=1}^{n-1} y_i \right)^{\alpha_n - 1} \cdot$$

$$\cdot I_S(y_1, \ldots, y_{n-1})$$

en donde S es el símplex

$$S = \left\{ (y_1, \ldots, y_{n-1}) : y_i \geq 0,\ i = 1, \ldots, n-1,\ \sum_{i=1}^{n-1} y_i \leq 1 \right\}$$

Para las principales propiedades de esta distribución, puede consultarse Wilks (1962).

Definición 1.2: Ferguson (1973)

Sea $(\mathcal{X}, \mathcal{A})$ un espacio medible y sea α una medida finita no nula sobre $(\mathcal{X}, \mathcal{A})$. Definimos un proceso estocástico particular $\{P(A) : A \in \mathcal{A}\}$ de la siguiente manera: P es un proceso de Dirichlet sobre $(\mathcal{X}, \mathcal{A})$ con parámetro α, si para cada $k = 1, 2, \ldots$ y para cada partición medible $(B_1, \ldots, B_k)$ de $\mathcal{X}$, la distribución de la variable aleatoria k-dimensional $(P(B_1), \ldots, P(B_k))$ es de Dirichlet con parámetros $(\alpha(B_1), \ldots, \alpha(B_k))$.

Si $[0, 1]^{\mathcal{A}}$ representa el espacio de todas las funciones de $\mathcal{A}$ en $[0, 1]$, este espacio incluye al conjunto de todas las medidas de probabilidad sobre $(\mathcal{X}, \mathcal{A})$, al cual denotaremos por $\mathcal{P} = \{P\}$. Sea $\beta_{[0,1]^{\mathcal{A}}}$ la σ-álgebra que contiene a todos los cilindros finito-dimensionales de $[0, 1]^{\mathcal{A}}$ y también a $\mathcal{P}$; Ferguson (1973) demuestra que existe una medida de probabilidad Q_α sobre $([0, 1]^{\mathcal{A}}, \beta_{[0,1]^{\mathcal{A}}})$ que admite a las anteriores distribuciones de Dirichlet como marginales finito-dimensionales y para la que $Q_\alpha(\mathcal{P}) = 1$. Puede demostrarse además que la probabilidad aleatoria P seleccionada por Q_α es, con probabilidad uno, discreta, Berk y Savage (1977).

Definición 1.3: Ferguson (1973)

La colección de variables aleatorias $X_1, \ldots, X_n$, con valores sobre $(\mathcal{X}, \mathcal{A})$, se dice que es una muestra de tamaño n obtenida a través de un proceso de Dirichlet P sobre $(\mathcal{X}, \mathcal{A})$ con parámetro α, si para todo $m = 1, 2, \ldots$ y conjuntos $\mathcal{A}$-medibles $A_1, \ldots, A_m$, $C_1, \ldots, C_n$, es

$$\Pr\{X_1 \in C_1, \ldots, X_n \in C_n / P(A_1), \ldots, P(A_m), P(C_1), \ldots, P(C_n)\} = \prod_{j=1}^{n} P(C_j)$$

en donde Pr denota probabilidad.

Intuitivamente, $X_1, \ldots, X_n$ es una muestra de tamaño n obtenida a través de un proceso de Dirichlet, si P es seleccionada aleatoriamente con la probabilidad Q_α y luego, dado P, $X_1, \ldots, X_n$ es una muestra obtenida con la probabilidad P.

Entre las propiedades que utilizaremos de los procesos de Dirichlet, Ferguson (1973), destacan:

a-1) Si X es una variable aleatoria con espacio muestral $(\mathcal{X}, \mathcal{A})$ y con probabilidad aleatoria P, la cual es un proceso de Dirichlet de parámetro α,

$$\Pr\{X \in A / P(A)\} = P(A)$$

a-2) En las mismas condiciones del apartado anterior,

$$\Pr\{X \in A\} = \frac{\alpha(A)}{\alpha(\mathcal{X})}$$

a-3) También se verifica que

$$E[P(A)] = \frac{\alpha(A)}{\alpha(\mathcal{X})}$$

En el caso particular de que $(\mathcal{X}, \mathcal{A})$ sea $(\mathbb{R}, \beta)$, tiene sentido hablar de «Función de distribución aleatoria, la cual es un proceso de Dirichlet de parámetro α», sin más que considerar subconjuntos β-medibles que sean de la forma $(-\infty, x]$.

2. MODELO DE ERROR DE MEDIDA

Sea θ una variable aleatoria con distribución $G(\theta)$ sobre un espacio paramétrico medible $(\Theta, \sigma_{\Theta})$, en donde Θ está contenido en un intervalo acotado de $\mathbb{R}$.

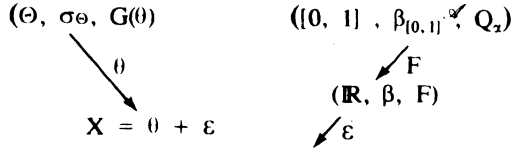
Sea ε una variable aleatoria real con distribución aleatoria F, la cual es un proceso de Dirichlet de parámetro α, y que representaremos por $F \in D(\alpha)$. Supondremos que α es una medida finita no nula sobre $(\mathbb{R}, \beta)$, que es absolutamente continua con respecto a la medida de Lebesgue en $\mathbb{R}$ y que tiene por densidad $\frac{d\alpha}{dl} = M h$, siendo $M > 0$ y h una densidad en $\mathbb{R}$.

El objetivo es estimar θ cuando observamos θ distorsionada por el ruido aditivo ε. Consideraremos una función de pérdida cuadrática.

Nuestro modelo será de la forma:

$$X = \theta + \varepsilon$$

en donde X es la variable aleatoria observable. Esquemáticamente tendríamos:



Nuestro objetivo es, pues, encontrar el estimador Bayes de θ, observado X. Al situarnos en un problema de decisión con pérdida cuadrática, el estimador Bayes será la medida a posteriori del parámetro,

$$\hat{\theta}_B = E[\theta/X]$$

Es necesario, por tanto, encontrar la distribución a posteriori de θ dado X.

Distribución a posteriori de θ dado X:

Sea $K_{\theta,F}(x)$ la distribución de X dados θ y F; tenemos que

$$K_{\theta,F}(x) = P_{\theta,F}\{X \leq x\} = P_{\theta,F}\{\varepsilon + \theta \leq x\} = F(x - \theta)$$

es decir, $K_{\theta,F}$ es un proceso de Dirichlet de parámetro α_θ, en donde

$$\alpha_\theta(-\infty, x] = \alpha(-\infty, x - \theta] \qquad [1]$$

Para ver la expresión [1] es suficiente con recordar que por la propiedad a-3) es

$$E[K_{\theta,F}(x)] = \frac{\alpha_\theta(-\infty, x]}{\alpha_\theta(\mathbb{R})}$$

y que por otro lado es

$$E[K_{\theta,F}(x)] = E[F(x - \theta)] = \frac{\alpha(-\infty, x - \theta]}{\alpha(\mathbb{R})}$$

Por otra parte, la distribución condicionada de X dado θ marginal de F será, en virtud de a-2), igual a

$$K_\theta(x) = P_\theta\{X \leq x\} = \frac{\alpha_\theta(-\infty, x]}{\alpha_\theta(\mathbb{R})} = \frac{\alpha(-\infty, x - \theta]}{\alpha(\mathbb{R})} \qquad [2]$$

y tendrá una densidad respecto a la medida de Lebesgue en $\mathbb{R}$, que será $h(x - \theta)$, siendo $M \cdot h$ la densidad de α en $\mathbb{R}$.

Por último, la densidad a posteriori de θ dado X se obtiene por la fórmula de Bayes y será:

$$\frac{d(G(\theta)/X)}{dG(\theta)} = \frac{h(x - \theta)}{\int_\Theta h(x - \theta) dG(\theta)} \qquad [3]$$

con lo que el estimador Bayes de θ será:

$$\hat{\theta}_B = \frac{\int_{\Theta} \theta h(x - \theta) dG(\theta)}{\int_{\Theta} h(x - \theta) dG(\theta)} \qquad [4]$$

3. PROCEDIMIENTOS EMPIRICO BAYES

Si consideramos el caso en el que $G(\theta)$ sea desconocida y en el que la medida α (parámetro del proceso) sea tal que su densidad respecto a la medida de Lebesgue en $\mathbb{R}$ sea de la forma

$$h(x/\theta) = h(x - \theta) = \theta^x l(\theta) m(x)$$

clase de funciones que bajo ciertas transformaciones pertenecen a la familia exponencial, el estimador Bayes será:

$$\hat{\theta}_B = \frac{\int_{\Theta} \theta^{x+1} l(\theta) m(x) dG(\theta)}{\int_{\Theta} \theta^x l(\theta) m(x) dG(\theta)} = \frac{m(x)}{m(x+1)} \cdot \frac{\int_{\Theta} h(x+1/\theta) dG(\theta)}{\int_{\Theta} h(x/\theta) dG(\theta)} =$$

$$= \frac{m(x)}{m(x+1)} \frac{h_G(x+1)}{h_G(x)}$$

en donde $h_G(x)$ es la densidad marginal de la variable aleatoria X, que es desconocida por serlo $G(\theta)$.

Consideremos ahora una sucesión de n problemas idénticos e independientes al anterior. En cada uno de ellos tendremos la terna (X_i, θ_i, F_i), $i = 1, \ldots, n$; X_i, variable aleatoria observable; θ_i, parámetro variable aleatoria, y $F_i \in D(\alpha)$, distribución aleatoria. Nuestro objetivo es encontrar para el problema $n + 1$, basándonos en las observaciones $x_1, \ldots, x_n; x_{n+1}$, un estimador $\hat{\theta}_n(x_1, \ldots, x_n; x_{n+1})$ que sea asintóticamente óptimo, en el sentido de que el riesgo Bayes de este estimador empírico converja hacia el riesgo Bayes cuando n tiende a ∞.

Para la construcción de $\hat{\theta}_n$, observemos la forma de $\hat{\theta}_B$:

$$\hat{\theta}_B = \frac{m(x)}{m(x+1)} \frac{h_G(x+1)}{h_G(x)}$$

Pues bien, utilizaremos como estimador empírico

$$\hat{\theta}_n(x_1, \ldots, x_n, x) = \frac{m(x)}{m(x+1)} \frac{h_n(x+1)}{h_n(x)}$$

en donde $h_n(x)$ es el estimador no paramétrico de la función de densidad $h_G(x)$, propuesto por Parzen (1962),

$$h_n(x) = \frac{1}{n\,c(n)} \sum_{j=1}^{n} k\left(\frac{x - X_j}{c(n)}\right)$$

en donde $k(y)$, núcleo del estimador, cumple las siguientes hipótesis:

1. $k(y)$ es una función de densidad simétrica.
2. $\int_{-\infty}^{+\infty} k(y) y^2 dy = 1$
3. $\int_{-\infty}^{+\infty} k(y) y^m dy < \infty$, $\forall m$, con $m = 0, 1, 2, \ldots$

y en donde $c(n)$, ventana del estimador, cumple las siguientes hipótesis:

1. $\lim_{n \to \infty} c(n) = 0$
2. $\lim_{n \to \infty} n\,c(n) = \infty$

Bajo estas hipótesis, Van Ryzin (1970) demostró la convergencia casi segura de $h_n(x)$:

$$h_n(x_1, \ldots, x_n, x) \xrightarrow{\text{a. s.}} h_G(x)$$

Teorema:

El estimador empírico Bayes $\hat{\theta}_n(X_1, \ldots, X_n, X) = \dfrac{m(X)}{m(X+1)} \cdot \dfrac{h_n(X+1)}{h_n(X)}$ es un estimador asintóticamente óptimo, en el sentido de que el riesgo Bayes de $\hat{\theta}_n$ es tal que $\lim_{n \to \infty} R_n(\hat{\theta}_n, G) = R(\hat{\theta}_B, G)$, siendo $R(\hat{\theta}_B, G)$ el riesgo Bayes del estimador Bayes $\hat{\theta}_B$.

El riesgo Bayes del estimador Bayes $\hat{\theta}_B$ es

$R(\hat{\theta}_B, G) = \int_{\mathcal{X}} (\text{Varianza a posteriori})\,dx$

es decir,

$R(\hat{\theta}_B, G) = \int_{\mathcal{X}} [E(\theta^2/x) - (E(\theta/x))^2]\,dx$

Por otra parte,

$$R_n(\hat{\theta}_n, G) = \int_{\mathscr{X}} R^*(\theta_n, x)\,dx$$

en donde

$$R^*(\theta_n, G) = \int_{\mathscr{X}^n} \prod_{i=1}^{n} h_G(x_i)\,dx_i \int_{\Theta} (\theta - \hat{\theta}_n)^2 h(x/\theta)\,dG(\theta) =$$

$$= \int_{\mathscr{X}^n} \prod_{i=1}^{n} h_G(x_i)\,dx_i (E(\theta^2/X = x) - 2\hat{\theta}_n E(\theta/X = x) + \hat{\theta}_n^2) =$$

$$= \int_{\mathscr{X}^n} \hat{\theta}^2 \prod_{i=1}^{n} h_G(x_i)\,dx_i - 2E(\theta/X = x) \int_{\mathscr{X}^n} \hat{\theta}_n \prod_{i=1}^{n} h_G(x_i)\,dx_i + E(\theta^2/X = x)$$

Como $\hat{\theta}_n \xrightarrow{\text{a. s.}} \hat{\theta}_B \Rightarrow \hat{\theta}_n^2 \xrightarrow{\text{a. s.}} \hat{\theta}_B^2$ y por el teorema de Lebesgue de la convergencia dominada será:

$$\lim_{n\to\infty} \int_{\mathscr{X}} \int_{\mathscr{X}^n} \hat{\theta}^2 \prod_{i=1}^{n} h_G(x_i)\,dx_i\,dx = \int_{\mathscr{X}} \hat{\theta}_B^2\,dx = \int_{\mathscr{X}} [E(\theta/X = x)]^2\,dx$$

y también

$$\lim_{n\to\infty} \int_{\mathscr{X}} \int_{\mathscr{X}^n} \hat{\theta}_n \prod_{i=1}^{n} h_G(x_i)\,dx_i\,dx = \int_{\mathscr{X}} \hat{\theta}_B \cdot dx = \int_{\mathscr{X}} E(\theta/X = x)\,dx$$

de donde

$$\lim_{n\to\infty} R_n(\hat{\theta}_n, G) = \int_{\mathscr{X}} (E(\theta^2/X = x) - [E(\theta/X = x)]^2)\,dx = R(\hat{\theta}_B, G)$$

como queríamos demostrar.

BIBLIOGRAFIA

BERK, R. H., y SAVAGE, I. R.: «Dirichlet processes produce discrete measures: an elementary proof». *Contributions to Statistics* (1977).

FERGUSON, T.: «A Bayesian analysis of some nonparametric problems». *The Annals of Statistics*, 1, 209-230 (1973).

PARZEN, E.: «On estimation of a probability density function and mode». *The Annals Mathematical Statistics*, Vol. 33 (1962).

ROBBINS, H.: «The empirical Bayes approach to Statistical decision problems». *The Annals Mathematical Statistics*, Vol. 35 (1964).

VAN-RYZIN, J.: «On some nonparametric empirical Bayes multiple decisions problems». *Nonparametric techniques in Statistical Inference* (1970).

WILKS, S.: *Mathematical Statistics*. J. Wiley (1962).

RESUMEN

Se considera un modelo de error de medida como problema de decisión bayesiana no paramétrico. Las variables asociadas a este problema de decisión, bajo pérdida cuadrática, tienen distribuciones $G(\theta)$ y F, respectivamente, en donde $G(\theta)$ es desconocida y F es un proceso de Dirichlet de parámetro α. Se construye una regla de decisión empírico Bayes en el caso de que la densidad del parámetro pertenezca a una cierta familia dada, la cual es asintóticamente óptima. Se comienza con un apartado preliminar dedicado a los procesos de Dirichlet.

Palabras clave: Modelo de error de medida, Procesos de Dirichlet, Estimadores empírico Bayes, Decisión bayesiana no paramétrica.

SUMMARY

Measure error model is considered as a nonparametric Bayesian decision. The variables associated with this decision problem, under quadratic loss, have the distributions $G(\theta)$ and F respectively, where $G(\theta)$ is unknown and F is a Dirichlet Process with parameter α. An Empirical Bayes decision rule is found, when the density function of the parameter belongs to a given family, which is asymptotically the best. We begin with a preliminary section dedicated to the Dirichlet processes.

Key words: Measure error model, Dirichlet processes, Empirical Bayes estimates, Nonparametric Bayesian decision.

AMS Subject classification: 62C10