Tema

6

Inferencia Bayesiana para Modelos Multiparamétricos

Índice

1. Introducción

Aunque ya se ha visto en el tema anterior un modelo multiparamétrico como el caso de la normal con ambos parámetros desconocidos, trataremos en este tema un modelo particular de caso multiparamétrico basado en la distribución multinomial que nos servirá para el análisis Bayesiano de datos de las tablas de contingencias.

2. La distribución de Dirichlet

La distribución de Dirichlet viene a generalizar al caso n-dimensional a la distribución beta. Tiene muchas aplicaciones en Biología, Fiabilidad Industrial, Economía, y sobre todo en el correspondiente enfoque Bayesiano.

Definición 2.1. *Se dice que el vector aleatorio* $\mathbf{X} = (X_1, \cdots, X_n)$ *sigue una distribución de Dirichlet de parámetros* $(\alpha_1, \cdots, \alpha_n)$ *con* $\alpha_i > 0 \quad \forall i$ *(y se denota como* $\mathbf{X} \sim Di(\alpha_1, \cdots, \alpha_n)$*) si su función de densidad es de la forma:*

$$f(x_1, \cdots, x_n | \alpha_1, \cdots, \alpha_n) = \frac{\Gamma(\sum_{i=1}^{n} \alpha_i)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} x_1^{\alpha_1 - 1} \cdots x_n^{\alpha_n - 1},$$

donde $0 \leq x_i \leq 1$ *y* $\sum_{i=1}^{n} x_i = 1$.

Nota 1. *A veces se emplea la misma notación para referirse a la inversa de la constante de proporcionalidad de la distribución de Dirichlet:*

$$Di(\alpha_1, \cdots, \alpha_n) = \frac{\prod_{i=1}^{n} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{n} \alpha_i)}$$

Proposición 2.2.

1. $E(X_j) = \frac{\alpha_j}{\sum_{i=1}^{n} \alpha_i}, \quad var(X_j) = \frac{\alpha_j}{A(A+1)} - \frac{\alpha_j^2}{A^2(A+1)}$ *con* $A = \sum \alpha_i$.
2. $cov(X_j, X_k) = -\frac{\alpha_j \alpha_k}{A^2(A+1)}$.
3. $X_j \sim Be(\alpha_j, A - \alpha_j)$.

3. Inferencia para la distribución de Dirichlet

3.1. La función de verosimilitud

Sean $X_1, \cdots, X_m$ variables aleatorias que representan las frecuencias para m categorías, siendo el total de observaciones un valor fijo $N = \sum_{i=1}^{m} x_i$, y que las probabilidades para cada categoría vienen dadas por el vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)$ donde $\theta_i \in [0, 1]$ y $\sum_{i=1}^{m} \theta_i = 1$. Entonces la distribución de probabilidad de $(X_1, \cdots, X_m)$ viene dada por:

$$P(X_1 = x_1, \cdots, X_m = x_m | \boldsymbol{\theta}) = \frac{N!}{\prod_{i=1}^{m} x_i!} \prod_{i=1}^{m} \theta_i^{x_i}.$$

Nótese que para $m = 2$ se tendría el modelo binomial.

3.2. A priori de Jeffrey

Se tiene que

$$\ln f(\mathbf{x}|\boldsymbol{\theta}) = \ln N! - \sum_{i=1}^{m} \ln x_i! + \sum_{i=1}^{m} x_i \ln \theta_i.$$

Luego,

$$\frac{\partial \ln f(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i} = \frac{x_i}{\theta_i}; \quad \frac{\partial^2 \ln f(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i^2} = -\frac{x_i}{\theta_i^2} \quad \text{y} \quad \frac{\partial^2 \ln f(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = 0.$$

Por tanto,

$$E\left[-\frac{\partial^2 \ln f(\mathbf{X}|\boldsymbol{\theta})}{\partial \theta_i^2}\right] = E\left[\frac{X_i}{\theta_i^2}\right].$$

Y ahora, teniendo en cuenta que $X_i \sim Bi(N, \theta_i)$ se llega a que

$$E\left[-\frac{\partial^2 \ln f(\mathbf{X}|\boldsymbol{\theta})}{\partial \theta_i^2}\right] = \frac{N}{\theta_i}.$$

Esto implica que los elementos de la matriz de Información de Fisher vienen dados por

$$[I(\theta_1, \ldots, \theta_m)]_{ij} = \begin{cases} \frac{N}{\theta_i} & \text{si } i = j; \\ 0 & \text{si } i \neq j. \end{cases}$$

Consecuentemente,

$$\det\left[I(\theta_1, \ldots, \theta_m)\right] = \prod_{i=1}^{m} \frac{N}{\theta_i}.$$

En definitiva, la distribución a priori de Jeffrey será:

$$\pi(\boldsymbol{\theta}) \propto \prod_{i=1}^{m} \theta_i^{-1/2}, \text{ es decir } \boldsymbol{\theta} \sim Di(\frac{1}{2}, \ldots, \frac{1}{2}.)$$

3.3. Distribución a posteriori de una familia conjugada

Observando la función de verosimilitud, y desde el enfoque de las familias conjugadas, se tomará una distribución a priori de Dirichlet con los parámetros conocidos:

$$\boldsymbol{\theta} \sim Di(a_1, \cdots, a_m).$$

Por tanto, la distribución a posteriori será:

$$\pi(\boldsymbol{\theta}|(x_1, \cdots, x_m)) \propto \prod_{i=1}^{m} \theta_i^{x_i} \prod_{i=1}^{m} \theta_i^{a_i - 1},$$

es decir

$$\boxed{\boldsymbol{\theta}|\mathbf{x} \sim Di(x_1 + a_1, \cdots, x_m + a_m)}$$

4. Tablas de Contingencia

Sin pérdida de generalidad y para comodidad en la notación vamos a estudiar el caso de tablas 2×2. Es decir, se tienen dos variables aleatorias cualitativas X e Y cada una de ellas con dos modalidades. Se extrae una muestra aleatoria de tamaño n y se observan las frecuencias de cada cruce:

$$n_{ij} \equiv \text{ n}^{\underline{\text{o}}} \text{ de observaciones de } x_i \text{ con } y_j, \text{ para } i,j = 1,2;$$

$$\theta_{ij} = P(X = x_i, Y = y_j) \text{ para todo } i,j.$$

La cuestión a resolver sería ¿existe independencia entre X e Y? En la Inferencia no-Paramétrica se ha visto el Test Exacto de Fisher, pero ¿cómo se podría resolver este problema si además añadimos algún tipo de información en el modelo y que se refiera a los parámetros?

En este caso hay dos modelos diferentes a contrastar: modelo de independencia frente a modelo de no-independencia. Es decir,

$$H_0 : M_D \rightarrow \text{ No Independencia;}$$
$$H_1 : M_I \rightarrow \text{ Independencia.}$$

En este caso, el factor de Bayes se calculará como:

$$FB = \frac{P(\mathbf{x}|M_D)}{P(\mathbf{x}|M_I)}.$$

La tabla de contingencia vendría dada por

	y_1	y_2
x_1	θ_{11}	θ_{12}
x_2	θ_{21}	θ_{22}

siendo el espacio paramétrico:

$$\mathbf{\Theta} = \{(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}), \text{ con } \theta_{ij} > 0, \sum \theta_{ij} = 1\}.$$

La distribución a priori que vamos a usar es la uniforme sobre dicho espacio paramétrico:

$$\boldsymbol{\theta} \sim Di(1,1,1,1).$$

Caso 1: Modelo de No-Independencia (M_D).

Se tiene que

$$P(\mathbf{x}|M_D) = \int_{\boldsymbol{\Theta}} P(\mathbf{x}|\boldsymbol{\theta}, M_D) P(\boldsymbol{\theta}|M_D) d\boldsymbol{\theta}.$$

Ahora bien,

$$P(\mathbf{x}|\boldsymbol{\theta}, M_D) = \frac{n!}{\prod n_{ij!}} \prod_{i,j} \theta_{ij}^{n_{ij}} \text{ donde } \mathbf{x} = (n_{11}, n_{12}, n_{21}, n_{22}).$$

Por tanto,

$$\begin{aligned} P(\mathbf{x}|M_D) &= \int_{\boldsymbol{\Theta}} \frac{n!}{\prod n_{ij}!} \prod_{i,j} \theta_{ij}^{n_{ij}} \frac{1}{Di(1,1,1,1)} d\boldsymbol{\theta} \\ &= \frac{n!}{\prod n_{ij!}} \frac{Di(n_{11}+1, \cdots, n_{22}+1)}{Di(1,1,1,1)}. \end{aligned}$$

Caso 2: Modelo de Independencia (M_I).

Este nuevo modelo afecta en realidad al espacio paramétrico, y por tanto, al recinto de integración. La restricción a usar bajo independencia es

$$\theta_{ij} = \theta_{i+}\theta_{+j} \quad \forall i, j.$$

Con lo cual nos quedaría:

$$P(\mathbf{x}|\boldsymbol{\theta}, M_I) = \frac{n!}{\prod n_{ij!}} \theta_{1+}^{n_{1+}} \theta_{2+}^{n_{2+}} \theta_{+1}^{n_{+1}} \theta_{+2}^{n_{+2}},$$

donde $\theta_{1+} = \theta_{11} + \theta_{12}$ sería la marginal para la primera fila y así sucesivamente.

En este caso la a priori adoptaría la forma

$$P(\boldsymbol{\theta}_+|M_I) = \frac{1}{Di(1,1)} \frac{1}{Di(1,1)} I_{\boldsymbol{\Theta}_+}(\boldsymbol{\theta}_+),$$

donde

$$\boldsymbol{\Theta}_+ = \{\boldsymbol{\theta} : \theta_{1+} + \theta_{2+} = 1 \text{ y } \theta_{+1} + \theta_{+2} = 1\}.$$

Luego realizando un cálculo similar al caso anterior queda que:

$$P(\mathbf{x}|M_D) = \frac{n!}{\prod n_{ij!}} \frac{Di(n_{1+}+1, n_{2+}+1) Di(n_{+1}+1, n_{+2}+1)}{Di(1,1) Di(1,1)}.$$

A partir de aquí el cálculo del Factor de Bayes es inmediato.

Como principio de acuerdo, y siempre de manera subjetiva, vamos a adoptar los puntos de corte dados por Jeffrey en cuanto al cálculo del Factor de Bayes (ver Cuadro 1.)

Factor de Bayes FB_{01}			Interpretación
	$>$	100	Evidencia extrema para H_0
30	$-$	100	Evidencia muy fuerte para H_0
10	$-$	30	Evidencia fuerte para H_0
3	$-$	10	Evidencia moderada para H_0
1	$-$	3	Evidencia anecdótica para H_0
	1		Sin evidencia
1/3	$-$	1	Evidencia anecdótica para H_1
1/10	$-$	1/3	Evidencia moderada para H_1
1/30	$-$	1/10	Evidencia fuerte para H_1
1/100	$-$	1/30	Evidencia muy fuerte para H_1
	$<$	1/100	Evidencia extrema para H_1

Cuadro 1: Tabla de puntos de corte para FB de Jeffrey