CAPíTULO 1

Simulación Metropolis-Hastings de una regresión lineal con ruido ARIMA y restricciones de desigualdad lineal

1.1. Descripción

Sea el modelo de regresión lineal sparse estandarizada de rango completo con restricciones lineales y con ruido ARIMA

$$\begin{gathered}
Y = X \cdot \beta + Z \\
A \cdot \beta \leq a \\
\phi(B) \triangle (B) Z_t = \theta(B) E_t \\
E \sim N\left(0, \sigma^2 I\right) \\
Y, Z, E \in \mathbb{R}^m \\
X \in \mathbb{R}^{m \times n} \\
a \in \mathbb{R}^r \wedge A \in \mathbb{R}^{r \times n} \\
rank(X) = n \\
m > n > 0 \\
r \geq 0 \\
n, m, r \in \mathbb{N}
\end{gathered} \tag{1.1.1}$$

donde

- β son los parámetros a estimar de la regresión,
- n es el número de parámetros a estimar de la regresión,
- m es el número de datos de contraste del modelo y ha de ser mayor que el de variables,
- r es el número de restricciones de inecuaciones lineales que puede ser eventualmente nulo, en cuyo caso se trataría de una regresión sin restricciones,
- Y es la matriz de output estandarizado del modelo y es completamente conocida.
- X es la matriz de iputs del modelo y es completamente conocida y de rango completo.
- A es la matriz de coeficientes de restricción, que es conocida y también podría cambiar en cada simulación
- a es el vector de frontera de restricción que es igualmente conocida aunque no necesariamente fija.
- E son los residuos independientes del modelo cuya distribución normal se propone como hipótesis principal del mismo, con media nula y desviación conocida σ, al igual que los polinomios DIF $\triangle(B)$, AR $\phi(B)$ y MA $\theta(B)$, cada uno de los cuales puede ser eventualmente el polinomio 1, incluso todos a la vez en cuyo caso estaríamos ante el caso particular de regresión lineal simple, sin parte ARIMA.

1.2. Evaluación de la densidad condicionada

La evaluación de la densidad condicionada para un vector β factible es trivial y muy rápida de ejecutar, tal y como se hace ahora mismo en el método pseudo MH, pues su logaritmo es, salvo una constante, el mismo que el de los residuos estandarizados

$$e = \frac{1}{\sigma} E \tag{1.2.1}$$

obtenidos resolviendo la ecuación en diferencias

$$\phi(B)\triangle(B)z_t = \theta(B)e_t \tag{1.2.2}$$

o bien aplicando cualquier descomposición simétrica de la matriz de covarianzas del proceso ARMA al ruido diferenciado normalizado

$$w_t = \triangle(B)z_t \tag{1.2.3}$$

donde

$$z = \frac{1}{\sigma}Z = \frac{1}{\sigma}(Y - X\cdot\beta) \tag{1.2.4}$$

$$p(\beta) \propto -\frac{1}{2}e^T e \tag{1.2.5}$$

1.3. Función generatriz de candidatos

Para tener un método de Metropolis-Hastings sólo necesitamos una función generatriz de candidatos que, o bien sea simétrica o bien tenga una densidad de cálculo sencillo. Para que el método sea eficiente la tasa de rechazo no debe ser muy alta. Para que el método sea coherente y robusto debe ser capaz de explorar todo el área de soluciones plausibles, es decir factibles con densidad significativa.

Una función generatriz simétrica es muy facil de obtener si no hay restricciones, pero, si las hay, estas lo complican todo enormemente, pues la posición de un punto concreto determina entornos factibles de volúmenes que pueden ser muy distintos si el punto está en el interior, lo cual permite un entorno esférico amplio, que si está más o menos cerca de la frontera, en cuyo caso tendrá un entorno semiesférico, y no digamos si está próximo a una arista o un vértice más o menos agudo con entornos en forma de cuña que puede llegar a ser muy estrecha.

A continuación se propondrá una familia paramétrica de funciones generatrices fáciles de evaluar y generar que permitan refinar de alguna forma la tasa de rechazo de forma dinámica, durante el propio proceso de simulación.

Estas funciones devolverán un punto del entorno del punto actual buscando un punto con densidad uniforme en una hiperesfera de radio ρ contenida íntegramente en la región factible

$$\begin{array}{c} \beta' = \beta + \lambda\cdot v \\ A\cdot(\beta+\lambda\cdot v) \le a \\ 0 \le \lambda \le \rho \\ \|v\| = 1 \end{array} \tag{1.3.1}$$

El radio ρ debe ser por tanto menor o igual que la distancia de β a la frontera del politopo

$$\beta \in \mathcal{P}(A,a) = \{\beta \in \mathbb{R}^n \perp A\cdot\beta \le a\} \tag{1.3.2}$$

es decir, la menor de las distancias de β a cada uno de los hiperplanos definidos por cada inecuación

$$\rho \le \langle\beta,\mathcal{P}(A,a)\rangle = \underset{i=1\ldots r}{min}\left\{\frac{\left|\left(\sum_{j=1}^{n} A_{ij}\beta_j\right) - a_i\right|}{\sum_{j=1}^{n} A_{ij}^2}\right\} \tag{1.3.3}$$

Pero β' también debe ser un punto "cercano" a β, luego el radio debe ser también menor que cierto parámetro auxiliar g que definiría el concepto de "cercanía" en la familia de funciones generatrices, el cual se podría ir modificando de forma que mejorara la calidad de la cadena. Así pues el radio queda definido en función del punto actual y el parámetro de cercanía

$$\rho = \rho(g,\beta) = min\{g, \langle\beta,\mathcal{P}(A,a)\rangle\} \tag{1.3.4}$$

La densidad de candidaturas es por tanto una uniforme en una esfera, es decir, dependiente sólo del punto de origen β y no del punto de destino β'

$$Q(\beta, \beta') = \frac{3}{4\pi\rho^3(g, \beta)} \tag{1.3.5}$$

Obsérvese que efectivamente no es, en general, una función simétrica respecto al orden de los parámetros, puesto que

$$Q(\beta', \beta) = \frac{3}{4\pi\rho^3(g, \beta')} \tag{1.3.6}$$

aunque sí lo sería en el caso sin restricciones o si ambos puntos están bien metidos en el interior del politopo, a distancia mayor que g de su frontera, ya que entonces el radio sería g en cualquier caso, es decir, independiente del punto de origen.

En cada simulación se puede aumentar o disminuir el valor del parámetro auxiliar de cercanía g de forma que el ratio de aceptación no sea ni muy alto ni muy bajo. Según resultados demostrados teóricamente[1]este ratio oscila entre el $50\,\%$ para $n = 1$ y un $23.4\,\%$ cuando $n \to \infty$.

Para generar un candidato β' a partir del actual β con esta distribución se procederá como sigue

1. Primero se simulará un vector multinormal estandarizado

$$u \sim Normal(0, I) \in \mathbb{R}^n \tag{1.3.7}$$

2. Dividiendo ese vector por su norma euclídea se obtendrá una dirección unitaria uniforme $v \in \mathbb{R}^n$ en la frontera de la hiperesfera de radio 1 centrada en el origen

$$v = \frac{u}{\|u\|_2} \tag{1.3.8}$$

3. Luego se multiplica ese vector v por un radio h con densidad directamente proporcional a la hipersuperficie de dimensión $n-1$ de la hiperesfera correspondiente

$$f(h) = c \cdot h^{n-1} \forall h \in [0, \rho] \tag{1.3.9}$$

cuya función de distribución será

$$\begin{array}{c} F(h) = \int_0^h c \cdot t^{n-1} \mathbf{d}t = \left.\frac{c}{n} t^n\right]_0^h = \frac{c}{n} h^n \\ F(\rho) = \frac{c}{n}\rho^n = 1 \Rightarrow c = n\rho^{-n} \\ F(h) = \left(\frac{h}{\rho}\right)^n \end{array} \tag{1.3.10}$$

luego su inversa es

$$F^{-1}(p) = \rho p^{\frac{1}{n}} \tag{1.3.11}$$

y podemos generar h a partir de una uniforme así

$$\begin{array}{c} h = \rho p^{\frac{1}{n}} \\ p \sim U(0,1) \end{array} \tag{1.3.12}$$

4. Finalmente se suma el vector de desplazamiento al punto actual

$$\beta' = \beta + h \cdot v \tag{1.3.13}$$

Esta familia de generatrices cumple todas las condiciones para implementar un método de simulación Metropolis-Hastings con todas las garantías de convergencia, mezclado y capacidad de exploración.

Otra cosa distinta es la velocidad de convergencia y la capacidad de recuperación cuando se cae en regiones de poca densidad. Para mejorar este aspecto se podría cambiar la uniformidad

[1]`http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aoap/1034625254`Roberts, G.O.; Gelman, A.; Gilks, W.R. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms". Ann. Appl. Probab. 7 (1): 110–120. doi:10.1214/aoap/1034625254

por otra densidad más dirigida al objetivo pero igualmente restringida a hiperesferas factibles de radio controlado como se ha descrito anteriormente.

Una forma muy sencilla de mejorar el método de generación de candidatos es modificar la métrica del problema original normalizando sus componentes para que la forma esférica de los entornos elegidos no perjudique ni beneficie a ninguna dirección en particular

$$\begin{gathered}\dot{Y} = \dot{X} \cdot \dot{\beta} + \dot{Z} \\ \dot{Z} = \tfrac{1}{\sigma} \triangle(B) Z \\ \dot{X} = \tfrac{1}{\sigma} \triangle(B) X \cdot D_X^{-1} \\ \dot{\beta} = D_X \cdot \beta \\ \dot{A} = A \cdot D_X^{-1}\end{gathered} \tag{1.3.14}$$

dondeD_X es la matriz diagonal de las normas de las columnas de $\triangle(B)X$. Con esto se equilibran los pesos de las componentes y se independiza el valor relativo del parámetro auxiliar de cercanía g.