Inferenza statistica per catene di Markov

(Breve riassunto informale, a cura di Luca La Rocca e Lilla Di Scala)

Sia P una matrice di transizione sullo spazio finito E. Si vuole fare inferenza non parametrica su P a partire dalle catene $X^1, \ldots, X^m$ di Markov omogenee per tale matrice. Se per ogni j la catena X^j parte dallo stato $x_0^j \in E$ e se le catene X^j sono tra loro indipendenti, la log-verosimiglianza del modello può scriversi

$$ll(X|P) = \sum_{(x,y)\in E^2} N_{x,y} \log P(x,y)$$

dove

$$N_{x,y} = \sum_{j=1}^{m} \sum_{k=1}^{n_j} \mathbf{1}_{\{X_{k-1}^j = x,\, X_k^j = y\}}$$

sono le transizioni da x a y contate osservando la catena X^j fino al tempo n_j.

Interessa in particolare il caso in cui tutte le catene partano dallo stesso stato ricorrente x_0 e vengano osservate fino al rientro in tale stato. In tale caso si può equivalentemente pensare a un'unica catena X osservata fino al tempo (aleatorio) $n = \sum_{j=1}^{m} n_j$ di m-esimo rientro in x_0.

Dalle equazioni normali, sinteticamente scritte come $\frac{\partial}{\partial P} ll(X,P) = 0$, tenendo conto che la matrice P deve essere stocastica, ovvero che deve aversi $\sum_{y\in E} P(x,y) = 1$ per ogni $x \in E$, si trova lo stimatore di massima verosimiglianza $\hat{P}$ definito da

$$\hat{P}(x,y) = \frac{N_{x,y}}{N_x}$$

dove $N_x = \sum_{y\in E} N_{x,y}$ è il totale delle transizioni osservate in uscita da x.

Lo stimatore $\hat{P}$ è consistente, perché risulta

$$\frac{N_{x,y}}{n} \longrightarrow \pi(x)P(x,y) \qquad \frac{N_x}{n} \longrightarrow \pi(x)$$

in conseguenza della Legge dei Grandi Numeri (si è indicata con π la legge invariante di P). Inoltre lo stimatore $\hat{P}$ è asintoticamente normale, perché

$$\frac{1}{\sqrt{n}}(N_{x,y} - N_x P(x,y)) \longrightarrow \mathcal{N}(0, \pi(x)P(x,y)(1 - P(x,y)))$$

in conseguenza del Teorema Centrale del Limite per indici aleatori.

Definito $\Delta^* = \{(x,y) \in E^2 : P^*(x,y) > 0\}$, è possibile sottoporre a verifica l'ipotesi $P = P^*$ mediante la statistica

$$T = \sum_{(x,y)\in\Delta^*} \frac{(N_{x,y} - N_x P^*(x,y))^2}{N_x P^*(x,y)}$$

che risulta asintoticamente distribuita come una chi-quadro con $r - s$ gradi di libertà, dove $r = \#\Delta^*$ ed $s = \#E$.

Se X è una catena su E con matrice P, allora Y definita da

$$Y_k = (X_k, X_{k+1}) \qquad k = 0, 1, \ldots$$

è una catena su $\Delta = \{(x, y) \in E^2 : P(x, y) > 0\}$ con matrice Q definita da

$$Q((x, y), (v, z)) = \delta_{\{y\}}(v) P(v, z);$$

la catena Y verrà detta catena ausiliaria associata alla catena originale X.

Applicando alla catena ausiliaria la metodologia sopra illustrata per la catena originale, si può verificare l'ipotesi $P = P^*$ mediante la statistica

$$U = \sum_{(x,y,z)\in\Gamma^*} \frac{(N_{x,y,z} - N_{x,y}P^*(y, z))^2}{N_{x,y}P^*(y, z)}$$

dove $\Gamma^* = \{(x, y, z) \in E^3 : (x, y) \in \Delta^* \ \& \ (y, z) \in \Delta^*\}$ e

$$N_{x,y,z} = \sum_{k=2}^{n} \mathbf{1}_{\{X_{k-2}=x,\, X_{k-1}=y, X_k=z\}}$$

sono le sequenze (x, y, z) osservate sino al tempo n. Tale statistica risulta asintoticamente distribuita come una chi-quadro con $t - r$ gradi di libertà, dove $t = \#\Gamma^*$ ed $r = \#\Delta^*$.

Si può allora pensare di verificare la bontà del modello stimato verificando l'ipotesi $P = \hat{P}$ sulla catena ausiliaria. Ciò può farsi mediante la statistica U che in questo caso particolare ($P^* = \hat{P}$) è asintoticamente distribuita come una chi-quadro con $t - 2r + s$ gradi di libertà, in virtù del fatto che provengono dai dati (cioè sono stati stimati) esattamente $r - s$ parametri (la statistica T relativa alla catena originale è invece banalmente nulla in questo pur interessante caso).

Per una dimostrazione dei risultati utilizzati nella giustificazione della metodologia sopra brevemente esposta si rimanda senz'altro a Dacunha-Castelle e Duflo [5]. La bibliografia è completata da una panoramica di riferimenti che trattano variamente l'argomento.

Bibliografia

[1] T. W. Anderson and L. A. Goodman. Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 28:89–110, 1957.

[2] I. V. Basawa and B. L. S. Prakasa Rao. *Statistical Inference for Stochastic Processes*. Academic Press, 1980.

[3] N. U. Bhat. *Elements of Applied Stochastic Processes*. John Wiley & Sons, 1972.

[4] P. Billingsley. Statistical methods in Markov chains. *Annals of Mathematical Statistics*, 32:12–40, 1961.

[5] D. Dacunha-Castelle and M. Duflo. *Probability and Statistics*, volume II. Springer-Verlag, 1986.

[6] R. Ganesan and A. T. Sherman. Statistical techniques for language recognition: An introduction and guide for cryptanalysts. *Cryptologia*, 4:321–366, 1993.

[7] M. L. Menéndez, D. Morales, L. Pardo, and K. Zografos. Statistical inference for finite Markov chains based on divergences. *Statisitcs & Probability Letters*, 41:9–17, 1999.