Bayesian inference for differential equations

Mario Beraha

Bayesian Statistics prof. A. Guglielmi

January 29, 2017

1 Introduzione

L'utilizzo di equazioni differenziali per modellizzare sistemi chimici e biologici ha, al giorno d'oggi, una storia ben consolidata; tuttavia, solo recentemente si è iniziato a porsi, e conseguentemente risolvere, il problema legato all'incertezza sul modello.

In molte applicazioni, la modellizzazione del sistema in considerazione non è ben caratterizzata nè dal punto di vista della scelta del modello "corretto" nè per quanto riguarda i valori dei parametri dello stesso.

L'incertezza in entrambi questi aspetti deve essere presa in considerazione in maniera sistematica se si vogliono fare simulazioni, o predizioni, in modo da evitare di giungere a conclusioni non giustificate o ingiustificatamente troppo ottimiste data l'incertezza di base.

Come mostrato nelle prossime pagine, la metodologia dell'inferenza bayesiana fornisce un setting coerente per quantificare l'incertezza propagatasi nel modello.

2 Inferenza bayesiana

Si considerano i dati osservati $\mathcal{D}\{\boldsymbol{y}, \boldsymbol{t}\}$ $\boldsymbol{y} \in \mathbb{R}^N$, $t \in \mathbb{R}$, dove $\boldsymbol{y}$ è l'osservazione legata a un certo stimolo t, che può essere visto semplicemente come il tempo. L'obiettivo è di stabilire una relazione funzionale $\phi : \; t \mapsto \boldsymbol{y}$.

Per fare ciò, definiamo la classe dei modelli che verranno presi in considerazione $M = \{M_1, M_2 \dots M_k\}$ a cui è assegnata una prior (discreta) $\pi = \{\pi(M_j)\}$

Se l'obiettivo è quello di fare previsione su un possibile valore y_* dato uno specifico t_*, ossia calcolare $p(y_*|t_*, \mathcal{D})$, sia l'incertezza sui parametri che quella sul modello deve essere presa in considerazione.

É facile osservare che:

$$p(y_*|t_*, \mathcal{D}) = \sum_{k \in M} p(y_*|t_*, \mathcal{D}, M_k)\pi(M_k|\mathcal{D}) \tag{1}$$

Ossia che la distribuzione predittiva è esprimibile come la media delle predittive di ogni modello. Usando il teorema di Bayes per modelli dominati, si ha l'espressione della posterior

$$\pi(M_k|\mathcal{D}) = \frac{p(\mathcal{D}|M_k)\pi(M_k)}{\sum_{\in M} p(\mathcal{D}|M_l)\pi(M_l)} \tag{2}$$