

SINTESI.

Una classe importante di problemi probabilistici e statistici é quella della stima di caratteristiche relative ad un certo processo aleatorio.

Esistono svariate tecniche di stima dei parametri descrittivi di un processo, tra le quali nominiamo quella di massima verosimiglianza, massima pseudo-verosimiglianza, tecniche bayesiane.

In questa tesi proponiamo tecniche che intendono incorporare l'incertezza per le osservazioni, nell'inferenza bayesiana.

Cio' che ci interessa é la ricostruzione di un processo aleatorio X , avendo a disposizione delle osservazioni y , relative a tale processo, che chiameremo dati e che sono realizzazioni di un vettore aleatorio Y , il quale rappresenta una versione degradata, a causa della presenza di certi fattori di disturbo del processo X .

Determinare quali sono le configurazioni piú verosimili del processo X , note le osservazioni, non é un problema semplice.

Comunque la conoscenza della forma del disturbo che ha prodotto le osservazioni Y , anziché il processo X , la conoscenza a priori del processo X e altre quantità legate ai dati Y , permette di fare inferenza.

C'è spesso una vasta scelta di plausibili ricostruzioni del processo X , note le osservazioni e altre quantità relative ad esso.

Quindi dobbiamo scegliere, in modo coerente, una di queste ricostruzioni che costituirá la nostra approssimazione del processo X .

La ricostruzione plausibile del processo X deve riflettere sia la conoscenza dei dati, che la conoscenza a priori del processo stesso.

L'inferenza che piú si addice per la risoluzione di queste nostre esigenze é, quindi, quella bayesiana, anche se alcune importanti applicazioni, come ad esempio quella per l'agricoltura, sono state affrontate con l'inferenza classica.

Attraverso la struttura bayesiana, l'inferenza viene basata su una distribuzione di probabilità che chiameremo a posteriori proprio perché tiene conto di tutte le nostre esperienze antecedenti l'inizio delle osservazioni y .

Quasi sempre, nell'affrontare problematiche reali delle quali si vogliono considerare una grande quantità di aspetti e caratteristiche, questa distribuzione a posteriori é difficilmente campionabile con metodi alternativi a quelli proposti in questa tesi.

Inoltre la difficoltà del campionamento aumenta con le dimensioni del problema che generalmente sono molto grandi.

Una possibile soluzione a questo problema é quella di semplificare il modello che schematizza il processo X , fino a ottenere una formulazione che si possa effettivamente calcolare. É evidente che molte informazioni saranno perdute adottando questa procedura di semplificazione, di cui diamo un esempio unidimensionale nel campo reale \mathfrak{R} qui di seguito.

Se indichiamo momentaneamente con $p(a, b]$ la probabilità a posteriori del sottoinsieme $(a, b]$ del campo reale \mathfrak{R} , che vogliamo calcolare, si dimostra che, con opportune ipotesi, vale la seguente rappresentazione di p ;

$$p(a, b] = \frac{1}{2\pi} \int_{\mathfrak{R}} ((e^{-ita} - e^{-itb})h(t)/it) dt$$

$i^2 = -1$, e

$$h(t) = \sum_{k=0}^{n+1} (it)^k / k!$$

$$\int_{\mathfrak{R}} z^k p(dz) + r_n(t),$$

$\forall t \in \mathfrak{R} \quad n \in \mathbf{N}$

Possiamo ora scegliere l'intero $n \in \mathbf{N}$, in modo che $h(t)$ sia calcolabile con l'errore $r_n(t)$, $\forall t \in \mathfrak{R}$.

Poiché la quantità $\int_{\mathfrak{R}} z^k p(dz)$ si chiama *momento di ordine k* , $k \in \mathbf{N}$, chiameremo questo modo di procedere *metodo dei momenti*.

(Un possibile riferimento é [1]).

In questo metodo, oltre alla difficoltà conseguente al calcolo del primo integrale, si aggiunge quella del calcolo dei momenti.

I metodi che esporremo non hanno bisogno del calcolo dei momenti, perché agiscono direttamente sulla distribuzione.

Le tecniche bayesiane che esporremo sono chiamate Monte Carlo per catene di Markov (MCMC).

Esse inizialmente furono utilizzate nell'ambito della Meccanica Statistica, nei modelli che sintetizzavano le interazioni tra le particelle di grandi sistemi multidimensionali discreti.

In questa tesi, abbiamo approfondito il modello di Ising unidimensionale e bidimensionale che sintetizza l'orientamento degli spins di un materiale opportunamente discretizzato e abbiamo illustrato argomentazioni inerenti il confronto tra due dei piú noti algoritmi MCMC applicati ad un particolare modello di Ising per le immagini digitalizzate.

Da pochi anni queste tecniche vengono usate come metodi di simulazione di campi stocastici di Markov sia nel contesto bayesiano che nel contesto numerico e altresí classico.

Inoltre le aree di ricerca dei campi di applicazione di tali tecniche sono in via di sviluppo. Comunque tali tecniche ben si adattano a situazioni difficili di calcolo delle probabilità e di statistica.

Nel capitolo degli algoritmi, abbiamo illustrato le tecniche Monte Carlo per catene di Markov iniziando con una classe risolutiva per spazi prodotto, detta *full conditional* e ricavata dalla distribuzione di interesse che, come si é detto, é una certa distribuzione a posteriori.

Inoltre abbiamo illustrato una seconda classe di tali tecniche, nella parte degli algoritmi, che può essere risolutiva per spazi qualunque e che é detta generalizzata.

Le due suddette classi di tecniche Monte Carlo per catene di Markov sono costituite da tre tipi di algoritmi ciascuno, che furono introdotte da Hastings nel 1970, da Metropolis nel 1953 e da Gibbs 1902 e dalle quali hanno preso i nomi.

In seguito furono prese in considerazione e migliorate da molti altri studiosi tra cui Grenander U., Geman S. and Geman D., Higdon D., Besag J., Mengersen K., Winkler G., Swartz

T., Evans M., Frigessi A., ai quali abbiamo fatto riferimento, oltre che agli effettivi inventori degli algoritmi, nello sviluppo e nel commento relativo a questa tesi.

Illustriamo ora un breve esempio per mostrare la tecnica di campionamento Gibbs full conditional per una distribuzione a posteriori $\pi(x/y)$ su \mathfrak{R}^2 , per certi dati $y \in \mathfrak{R}^2$, che indichiamo per brevità con $\mu(x)$.

Indichiamo inoltre con $\mu|_{\mathfrak{R}}(x_1/x_2)$ la distribuzione $\pi|_{\mathfrak{R}}(x_1/x_2, y)$, ed inoltre con $\mu|_{\mathfrak{R}}(x_2/x_1)$ la distribuzione $\pi|_{\mathfrak{R}}(x_2/x_1, y)$ che rappresentano le full conditionals o caratteristiche locali di $\mu|_{\mathfrak{R}}(x)$, $\forall x \in \mathfrak{R}^2$.

Supponiamo per semplicità di disporre di un campione $x^{(0)}$ estratto dalla distribuzione μ e generiamo ulteriori campioni

$$x^{(1)}, x^{(2)}, \dots, x^{(n)},$$

$n \in \mathbb{N}$, estratti dalla distribuzione μ , attraverso campionamenti unidimensionali delle caratteristiche locali, con la seguente metodologia.

Abbiamo che $x^{(0)} = (x_1^{(0)}, x_2^{(0)})$ è un campione estratto dalla distribuzione μ .

Aggiorniamo, rimpiazzandola, la componente $x_2^{(0)}$ del vettore $x^{(0),1} = (x_1^{(0)}, x_2^{(0)})$, con un campione $x_2^{(1)}$ estratto dalla distribuzione $\mu|_{\mathfrak{R}}(x_2/x_1^{(0)})$.

Se poniamo $x^{(1)} = (x_1^{(1)}, x_2^{(1)})$, essendo $x^{(0)}$ un campione estratto dalla distribuzione μ , otteniamo che anche $x^{(0),1}$ è un campione estratto dalla distribuzione μ e $x^{(1)}$ è un campione estratto dalla distribuzione μ .

Continuando ad iterare $n - 1$ volte questa procedura, otteniamo dei campioni

$$x^{(2)}, x^{(3)}, \dots, x^{(n)}$$

estratti dalla distribuzione μ .

Poiché, inoltre, ogni $x^{(i)}$ dipende solo da $x^{(i-1)}$, $\forall i \in (0, 1, \dots, n)$, può essere interpretata come una realizzazione di una variabile aleatoria $X^{(i)}$ di una catena di Markov con distribuzione stazionaria μ .

L'aggiornamento di tutte le componenti del vettore corrente $x^{(i)}$ attraverso campionamenti unidimensionali delle caratteristiche locali, è un ingrediente fondamentale per tutte le tecniche Monte Carlo per catene di Markov full conditionals, che approfondiremo nel capitolo dedicato agli algoritmi.

Si noti, inoltre, la necessità di saper estrarre campioni unidimensionali, che abbiamo affrontato nel capitolo dedicato agli algoritmi di simulazione diretta, utili anche per estrarre campioni direttamente dalla distribuzione a posteriori $\pi(x/y)$, qualora essa abbia una formulazione particolarmente semplice, senza generare una catena di Markov.

Inoltre, per far fruttare questo schema, è necessario sapere se per un qualunque $x^{(0)}$ estratto da una distribuzione iniziale ν su \mathfrak{R}^2 , $x^{(i)}$ converge, quando $i \rightarrow \infty$, in qualche senso ad un campione x estratto dalla distribuzione μ e se un qualsiasi funzionale di π , che indichiamo con $f(\pi(x/y))$, può essere stimato con il funzionale $f(\pi_i^*(x/y))$, dove $\pi_i^*(x/y)$ è la distribuzione empirica a posteriori

$$i^{-1} \sum_{t=1}^i I_{(X^{(t)} \leq x)}$$

e si é indicato con I_E , la funzione indicatrice dell'insieme E, $i \in \mathbf{N}$.

In sintesi, una qualunque tecnica Monte Carlo per Markov, usa una realizzazione parziale $x^{(1)}, x^{(2)}, \dots, x^{(n)}$, $n \in \mathbf{N}$, di una opportuna catena di Markov, $X^{(1)}, X^{(2)}, \dots$, definita su un certo spazio W, munito di una σ algebra, e di una misura di probabilità σ finita su W, che ha come distribuzione stazionaria e limite, la distribuzione a posteriori che vogliamo analizzare e che indichiamo con $\pi(x/y)$.

Attraverso queste realizzazioni parziali, possiamo stimare un funzionale di π , che indichiamo con $f(\pi(x/y))$, con il funzionale $f(\pi_n^*(x/y))$, dove $\pi_n^*(x/y)$ é la distribuzione empirica a posteriori

$$n^{-1} \sum_{t=1}^n I_{(X^{(t)} \leq x)}$$

e si é indicato con I_E , la funzione indicatrice dell'insieme E, $n \in \mathbf{N}$.

Nel capitolo relativo ai teoremi limite, abbiamo discusso proprietà di convergenza delle catene di Markov che nel nostro caso sono omogenee.

Le catene di Markov omogenee, sono quelle tali che in un certo istante $t \in \mathbf{N}$, dipendono solo da ciò che é accaduto all'istante $t - 1$ e non ciò che é accaduto negli istanti $t - 2, t - 3, \dots, 1$ e tale dipendenza non é funzione del tempo.

Sempre nel capitolo relativo ai teoremi limite, abbiamo discusso proprietà di convergenza dei funzionali delle distribuzioni empiriche ai funzionali della distribuzione limite e stazionaria. Una importante applicazione di queste tecniche algoritmiche é il calcolo di integrali.

Non appena le dimensioni degli integrali crescono, le difficoltà di calcolo aumentano, sebbene anche per piccole dimensioni non sempre é possibile ottenere la soluzione esatta. Nella statistica abbiamo spesso bisogno di approssimare integrali del tipo

$$E(h) = \int_{\mathfrak{R}^2} g(x)h(x)dx$$

dove g é una funzione tale che $g: \mathfrak{R}^k \rightarrow \mathfrak{R}^+$, $g: x \rightarrow g(x)$ e h é una funzione tale che $h: \mathfrak{R}^k \rightarrow \mathfrak{R}$, $h: x \rightarrow h(x)$, che in genere sono rispettivamente il prodotto della distribuzione a priori con la distribuzione dei dati noti i parametri x e una funzione con media a posteriori da calcolare

$$E(h) = \int_{\mathfrak{R}^2} g(x)h(x)dx$$

che rappresenta un problema di integrazione ponendo $g = 1_E$, dove 1_E é la funzione indicatrice dell'insieme E.

Inoltre assumiamo che hg é una funzione integrabile su \mathfrak{R}^k .

Abbiamo trattato questo aspetto nel capitolo dei teoremi limite, nelle osservazioni sulla utilità della media empirica ottenuta attraverso realizzazioni di opportune catene di Markov. ■

Le applicazioni delle tecniche bayesiane sono svariate, e vanno dall'agricoltura, ecologia, morfologia, economia, alla medicina etc. . . . etc..

Quelle che abbiamo trattato nel capitolo relativo alle applicazioni sono: un approccio bayesiano all'analisi statistica di esperimenti sulle varietà di coltivazioni agricole, nella parte relativa all'agricoltura, un approccio bayesiano all'analisi delle immagini digitalizzate

di un esame medico, nella parte relativa alla tomografia, e un approccio bayesiano all'analisi statistica di immagini digitalizzate di una videocamera, nella parte relativa all'analisi delle immagini.

Osservazioni sulla utilità della media empirica ottenuta attraverso realizzazioni di opportune catene di Markov.