Progetto strategico CNR *Decisioni statistiche: teoria e applicazioni*
Pitagora Editrice, Bologna 1999.

INFERENZA BAYESIANA NON PARAMETRICA

PIETRO MULIERE e SONIA PETRONE
Università di Pavia e Università dell'Insubria

Riassunto.
In questa relazione vengono presentati i contributi dei componenti del progetto strategico CNR "*Decisioni statistiche: teoria e applicazioni*" sul tema dell'inferenza bayesiana non parametrica.

L'inferenza bayesiana si e' sviluppata tradizionalmente in ambito parametrico; tuttavia, a partire dalla fine degli anni 60, le procedure bayesiane non parametriche hanno avuto una rilevanza crescente, sia dal punto di vista teorico che computazionale. Ci pare di poter affermare che il contributo della scuola italiana sia stato decisamente rilevante per questo sviluppo.

In particolare, nella presente relazione vengono discussi alcuni problemi di base nell'inferenza bayesiana nonparametrica: la scelta fra impostazione parametrica o non parametrica; l'assegnazione di una distribuzione iniziale su una misura di probabilità aleatoria; la determinazione della distribuzione finale e della distribuzione di funzionali; l'analisi nonparametrica di dati con strutture di dipendenza complesse (in particolare la scambiabilità parziale).

Il paragrafo finale della relazione contiene alcuni riferimenti storici e bibliografici.

1 Introduzione.

Questa relazione ha lo scopo di presentare i contributi dei componenti del progetto strategico CNR "*Decisioni statistiche: teoria e applicazioni*", 1998, sul tema dell'inferenza bayesiana non parametrica. Il taglio scelto è il seguente: ci proponiamo di individuare alcuni *problemi generali* che si pongono, a nostro avviso, nella ricerca in ambito bayesiano non parametrico; quindi vedremo quali contributi, problema per problema, siano stati forniti dai componenti del progetto. La trattazione sarà relativamente informale; le proposte saranno delineate nelle linee essenziali, rinviando ai lavori degli autori per i dettagli formali.

L'inferenza bayesiana si e' sviluppata tradizionalmente in ambito parametrico; tuttavia, a partire dalla fine degli anni 60, le procedure bayesiane non parametriche hanno avuto una rilevanza crescente, sia dal punto di vista teorico che computazionale (si veda il paragrafo 6 di questa relazione). Ci pare di poter affermare che il contributo della scuola italiana sia stato decisamente rilevante per questo sviluppo.

I problemi di base che si pongono nell'inferenza bayesiana non parametrica, e che saranno discussi in questa relazione, sono i seguenti:

* Un problema preliminare: parametrico o non parametrico?
* Assegnazione di una distribuzione iniziale su una misura di probabilità aleatoria.
* Determinazione della distribuzione finale e della distribuzione di funzionali.
* Strutture di dipendenza più complesse: scambiabilità parziale.

Nel paragrafo finale della relazione daremo alcuni riferimenti storici e bibliografici.

2 Un problema preliminare: parametrico o nonparametrico?

Si possono dare molte motivazioni a favore o sfavore dell'approccio parametrico o non parametrico ai problemi inferenziali, ma una motivazione di base e' fornita da Regazzini [11], con riferimento a osservazioni $X_1, X_2, \ldots$ scambiabili. La scambiabilità e' lo schema di base nell'impostazione bayesiana ed esprime la situazione di esperimenti ripetuti in condizioni omogenee; la scambiabilità e' pertanto l'ipotesi di base adottata nelle prime 3 sezioni di questa relazione.

Se $X = \{X_1, X_2, \ldots\}$ e' una successione di variabili aleatorie a valori in $\mathcal{X} \subseteq \Re^d$ e supponiamo che la legge di probabilità P di X sia scambiabile, allora, per il teorema di rappresentazione di de Finetti, per ogni $n = 1, 2, \ldots$ e per ogni $(x_1, \ldots, x_n)$ si ha

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{F}} \prod_{i=1}^{n} F(x_i) d\pi(F) . \qquad (1)$$

Inoltre la successione delle funzioni di ripartizione empiriche $F_n(x) = \frac{\sum_{i=1}^{n} \delta_{(-\infty, x]}(x_i)}{n}$ (dove $\delta_A(\cdot)$ e' la funzione indicatrice dell'insieme A) converge debolmente a $\tilde{F}$, quasi certamente rispetto a P. La legge di probabilità π e' la legge di probabilità di $\tilde{F}$ (misura di de Finetti) ed e' univocamente determinata. Dunque dall'ipotesi di scambiabilità segue la possibilità di rappresentare la distribuzione di X per mezzo di un *modello statistico*, che e' la realizzazione di $\tilde{F}$ in corrispondenza a $x = \{x_1, x_2, \ldots\}$, ed una *misura di probabilità iniziale* π.

La formulazione nonparametrica (1) data sopra e' la più appropriata per cogliere il significato del teorema di rappresentazione di de Finetti; ovvero, se assumiamo solo la scambiabilità la rappresentazione coinvolge un parametro infinito-dimensionale che e' il limite debole della successione delle funzioni di ripartizione empiriche, con probabilità uno. Per giustificare la dipendenza di tale limite da un parametro finito-dimensionale, e quindi un approccio *parametrico*, devono essere introdotte ulteriori ipotesi circa gli osservabili.

Tali ipotesi aggiuntive possono portare a restringere la classe delle possibili realizzazioni-limite della funzione di ripartizione empirica (che nella (1) e' la classe di tutte le funzioni di ripartizione sullo spazio campionario) ad una sottoclasse $\mathcal{F}_\Theta \subset \mathcal{F}$ le cui f.r. sono indicizzate da un parametro finito dimensionale θ. Quali informazioni aggiuntive sono rilevanti per scegliere un modello F_θ, o in altre parole per ridurre il parametro infinito-dimensionale ad un parametro finito-dimensionale?

Il problema e' evidentemente amplissimo e molto dibattuto nella letteratura recente, che sovente si propone di ricercare meccanismi automatici per la scelta di un modello statistico (parametrico). Nei contributi del gruppo di ricerca (Regazzini e Sazonov [12] e [13]; Fortini, Ladelli e Regazzini [2]) si possono individuare due interessanti linee di ricerca relativamente a questo problema.

Regazzini e Sazonov ([12] e [13]) suggeriscono di sostituire il parametro infinito-dimensionale con le frequenze limite corrispondenti ad una partizione finita dello spazio campionario, riducendo così il problema ad un problema con parametro finito-dimensionale. La proposta di Regazzini e Sazonov sarà discussa nel paragrafo 3 relativo alla assegnazione di una distribuzione iniziale in ambito nonparametrico; infatti, si tratta in realtà della proposta di una distribuzione iniziale molto generale, che consente di approssimare, con margine di approssimazione arbitrario, una qualunque distribuzione iniziale, dunque in particolare una qualunque distribuzione iniziale in ambito nonparametrico.

In effetti, pur ottenendo la riduzione del problema ad un problema con parametro finito-dimensionale, la dimensione del parametro nella proposta di Regazzini e Sazonov può essere ancora molto elevata. Una proposta che consente invece di ottenere un parametro di dimensione finita e *piccola* e' suggerita da Fortini, Ladelli e Regazzini [2] e Regazzini [11] a partire da una visione previsiva dell'inferenza. Essi mostrano che, qualora si ritenga sufficiente conoscere una sintesi delle osservazioni passate $(X_1 = x_1, \ldots, X_n = x_n)$ ai fini della previsione dell'osservazione successiva X_{n+1} (riassunto esaustivo a fini previsivi), allora e' possibile rappresentare la legge scambiabile P di X per mezzo di un modello parametrico in cui la verosimiglianza e' il limite della successione delle distribuzioni predittive e la legge di probabilità del parametro e' la legge limite della successione dei riassunti esaustivi a fini previsivi. Le condizioni poste da Fortini Ladelli e Regazzini per ottenere tale risultato sono meno restrittive di quelle utilizzate in precedenti risultati di questo tipo (Campanino e Spizzichino, 1981; Cifarelli e Regazzini, 1980 e 1981).

Dunque, come si e' detto inizialmente, la sola ipotesi di scambiabilità porta ad una rappresentazione della legge di probabilità di X di tipo nonparametrico. Tuttavia, se, in una visione previsiva dell'inferenza, riteniamo possibile assumere l'esistenza di riassunti esaustivi a fini previsivi, possiamo ottenere una rappresentazione per mezzo di un modello parametrico, i cui parametri hanno significato come limiti di osservabili.

3 Assegnazione di una distribuzione iniziale su una misura di probabilità aleatoria.

Secondo de Finetti, la previsione e' in generale lo scopo primario dell'inferenza statistica, ed e' risolta nella specificazione delle distribuzioni di probabilità predittive di $X_{n+1} \mid X_1 = x_1, \ldots, X_n = x_n, n = 1, 2, \ldots$. Tuttavia, poichè la specificazione *diretta* delle distribuzioni predittive e' in generale difficile, nell'ipotesi di scambiabilità e' possibile utilizzare il teorema di rappresentazione come un passo intermedio per quantificare opinioni iniziali sugli osservabili; nella rappresentazione, la distribuzione iniziale π può essere interpretata come un'approssimazione della legge di probabilità della f.r. empirica (che e' osservabile), per n sufficientemente grande. Utilizzando la terminologia in Regazzini [11], diciamo allora *diretta* ogni inferenza sugli osservabili X, e *indiretta* ogni inferenza sui modelli statistici. L'inferenza diretta corrisponde all'inferenza predittiva; l'inferenza indiretta comprende l'analisi parametrica e non parametrica.

Nella letteratura bayesiana, l'inferenza indiretta e' sicuramente quella più ampiamente trattata, e risolta mediante la formulazione di una distribuzione iniziale e il calcolo della distribuzione finale mediante l'algoritmo bayesiano. Se adottiamo una visione nonparametrica, il problema di partenza e' allora come assegnare una distribuzione iniziale che abbia come supporto la classe di tutte le misure di probabilità sullo spazio campionario. Questo e' il problema analizzato in questo paragrafo.

Come dicevamo, l'approccio inferenziale indiretto e' quello comunemente adottato in letteratura, anche in problemi in cui lo scopo ultimo dell'analisi e' la previsione. Ci sembra di poter affermare che i contributi del gruppo di ricerca abbiano un ruolo essenziale per sottolineare invece l'importanza di una visione diretta dell'inferenza. Anche relativamente al problema della assegnazione di una distribuzione iniziale in ambito nonparametrico, la visione diretta risulta molto fruttuosa, in quanto e' possibile caratterizzare leggi di probabilità iniziali nonparametriche a partire da ipotesi sulle distribuzioni predittive.

In questo paragrafo illustreremo alcuni contributi del gruppo di ricerca relativamente al tema della scelta di una iniziale non parametrica, suddividendo la presentazione secondo i seguenti punti:

(a) caratterizzazione della distribuzione iniziale mediante ipotesi predittive;

(b) approssimazione dell'iniziale mediante misture finite di distribuzioni di Dirichlet;

(c) costruzione della distribuzione iniziale mediante polinomi.

(a) *Caratterizzazione della distribuzione iniziale mediante ipotesi predittive.*

Sia $X = \{X_n, n = 1, 2, \ldots\}$ una successione infinita di variabili aleatorie scambiabili. La legge di probabilità della successione X può essere assegnata, secondo l'impostazione completamente predittiva, specificando la distribuzione P_1 di X_1 e le distribuzioni predittive P_n di $X_{n+1} \mid X_1 = x_1, \ldots, X_n = x_n$, per $n = 1, 2, \ldots$. Mentre il teorema di estensione di Ionescu-Tulcea fornisce condizioni di consistenza che garantiscono l'esistenza di un'unica legge di probabilità per $\{X_n, n = 1, 2, \ldots\}$, determinata dalla successione delle predittive P_n, Fortini, Ladelli e Regazzini ([2]; si veda anche Regazzini, [11]) forniscono condizioni necessarie e sufficienti sulla successione P_n affinchè esse caratterizzino una legge P scambiabile per la successione X. Questo risultato fornisce una caratterizzazione della scambiabilità in termini puramente predittivi.

Una volta assegnata la legge P scambiabile di X tramite la successione delle predittive, la misura di de Finetti di X risulta univocamente determinata mediante il teorema di rappresentazione di de Finetti. Questa impostazione e' stata utilizzata per caratterizzare alcune distribuzioni iniziali note in letteratura: si vedano Regazzini (1978) e Lo (1991) per il processo di Dirichlet, Walker e Muliere (1997*a*) per i Polya trees, Walker e Muliere (1997*b*) per il processo beta-Stacy.

(b) *Approssimazione dell'iniziale mediante misture finite di distribuzioni di Dirichlet.*

Regazzini e Sazonov ([12] e [13]) mostrano che, nel caso di osservazioni scambiabili a valori in uno spazio polacco, ogni distribuzione iniziale può essere approssimata (nel senso della metrica di Prokhorov), con grado di precisione arbitrario, mediante una mistura finita di distribuzioni di Dirichlet. Il risultato di Regazzini e Sazonov e' costruttivo nel senso che la mistura approssimante viene esplicitamente costruita. Alcuni risultati in questa direzione si trovano in Dalal (1978), Dalal e Hall (1980 e 1983) e Diaconis e Ylvisaker (1985).

Il lavoro di Regazzini e Sazonov e' suddiviso in due parti: nella prima si restringe l'attenzione ad uno spazio campionario $\mathcal{X}$ finito; la seconda generalizza ad uno spazio polacco $\mathcal{X}$. Forniamo una breve descrizione della costruzione di Regazzini e Sazonov nel caso di $\mathcal{X}$ finito, indicando le idee essenziali che portano alla generalizzazione a spazi campionari più generali. Ove non crei confusione, indicheremo con lo stesso simbolo la misura di probabilità e la corrispondente funzione di ripartizione.

Spazio campionario finito. Consideriamo una successione $X = \{X_1, X_2, \ldots\}$ di numeri aleatori scambiabili, a valori in uno spazio $\mathcal{X}$ finito

$$\mathcal{X} = \{a_1, \ldots, a_k, a_{k+1}\}.$$

Il teorema di rappresentazione di de Finetti consente in questo caso di rappresentare la legge di probabilità P di X come

$$P(X_1 = x_1, \ldots, X_n = x_n) = \int_{S_k} \theta_1^{n_1} \cdots \theta_k^{n_k} (1 - \sum_{i=1}^{k} \theta_i)^{n_{k+1}} \pi(d\theta) \tag{2}$$

dove S_k e' il simplesso k-dimensionale, $n_j = \{$numero di $(x_1, \ldots, x_n)$ uguali a $a_j\}$. La (2) e' una mistura di distribuzioni multinomiali, e π e' l'iniziale per il modello multinomiale. L'iniziale coniugata al modello multinomiale e' la distribuzione di Dirichlet; tale scelta e' ovviamente comoda, ma potrebbe non rappresentare le opinioni iniziali del ricercatore; tuttavia, la scelta di una diversa iniziale comporta in generale un maggiore onere computazionale. Regazzini e Sazonov [12] mostrano che ogni iniziale π per il modello multinomiale può essere approssimata da una opportuna mistura finita di distribuzioni di Dirichlet. Consideriamo dapprima una copertura del simplesso S_k mediante "piccoli ipercubi" di lati $(\frac{\nu_i - 1}{N}, \frac{\nu_i}{N}]$, dove $\nu = 0, 1, \ldots, N$ e $\sum \nu_i \leq N + k - 1$ e valutiamo la probabilità iniziale $\pi(\Delta_\nu)$ di ogni piccolo ipercubo Δ_ν. Consideriamo quindi una mistura di distribuzioni di Dirichlet in cui i pesi corrispondono ai valori $\pi(\Delta_\nu)$:

$$\sum_{\nu} \pi(\Delta_\nu) D(\cdot; \mu(\nu_1), \ldots, \mu(\nu_k), \lambda - \sum_{i=1}^{k} (\mu(\nu_i))), \tag{3}$$

dove $D(\cdot; \alpha_1, \ldots, \alpha_{k+1})$ indica la distribuzione di Dirichlet k-dimensionale, di parametri $(\alpha_1, \ldots, \alpha_{k+1})$. Per una determinata scelta dei parametri λ e μ (si veda per esempio Altomare e Campiti, 1994), la (3) corrisponde al polinomio di Bernstein sul simplesso S_k che approssima la funzione di ripartizione π. Regazzini e Sazonov trovano una espressione più generale dei parametri λ e μ, che garantisce che la mistura (3) sia vicina (nel senso della metrica di Prokhorov) alla distribuzione iniziale π. Forniscono inoltre il grado di approssimazione di π.

Caso generale. Questa costruzione può essere estesa al caso generale di spazio campionario $\mathcal{X}$ polacco. Ci restringiamo qui, per brevità , al caso in cui $\mathcal{X} \subseteq \Re^d$, in modo da avere il teorema di rappresentazione nella forma (1). Indichiamo con $\tilde{F}$ la misura di probabilità aleatoria (modello statistico) che ha legge di probabilità π (iniziale). Diamo solo l'idea generale della costruzione di Regazzini e Sazonov in questo contesto, che si basa su una "discretizzazione" di $\mathcal{X}$.

* Discretizziamo lo spazio campionario, mediante opportuni punti $a_1, \ldots, a_k$, quindi consideriamo una corrispondente "discretizzazione" $\tilde{F}_\epsilon$ del modello $\tilde{F}$. Sia π la distribuzione iniziale di $\tilde{F}$ e π_ϵ l'iniziale di $\tilde{F}_\epsilon$; e' possibile fare in modo che $d(\pi, \pi_\epsilon) < \epsilon$, dove $d(\cdot)$ indica la distanza di Prokhorov;

* in tal modo ci riconduciamo al caso di $\mathcal{X}$ finito e abbiamo la rappresentazione della legge di X mediante un modello multinomiale e una distribuzione iniziale, diciamo G_π;

* per i risultati precedenti, G_π può essere approssimata mediante una mistura finita di distribuzioni di Dirichlet, con grado di approssimazione arbitrario; si può anche mostrare che ad essa corrisponde una data mistura di distribuzioni di Dirichlet, diciamo $\pi_{\lambda,\mu}$, che approssima π_ϵ con precisione arbitraria;

* infine, si può mostrare che

$$d(\pi, \pi_{\lambda,\mu}) \leq d(\pi, \pi_\epsilon) + d(\pi_\epsilon, \pi_{\lambda,\mu})$$

e poiche' il membro di destra può essere piccolo a piacere, si ha che l'iniziale π può essere approssimata (nella metrica di Prokhorov) con una mistura di Dirichet esplicitamente costruita, con precisione arbitraria.

In effetti, il concetto di distribuzioni iniziali "vicine" può anche essere inteso come "vicinanza" delle corrispondenti distribuzioni finali, o delle distribuzioni predittive (si veda per esempio in Dalal e Hall, 1983; Krasker, 1984; Andreev e Arjas, 1996). Regazzini e Sazonov mostrano anche che, se le misure di probabilità iniziali $\pi_{\lambda,\mu}$, costruite secondo lo schema da essi proposto, convergono debolmente ad una iniziale π, allora la successione delle distribuzioni finali dedotte dalle $\pi_{\lambda,\mu}$ converge debolmente alla finale dedotta da π (fornendo una stima dell'errore di approssimazione, sotto ulteriori ipotesi). Tali risultati forniscono un metodo per esprimere opinioni iniziali in ambito bayesiano nonparametrico e per approssimare la distribuzione predittiva (nel senso della distanza variazionale) e la distribuzione finale di funzionali del tipo $\int \psi d\tilde{F}$ (nella metrica di Levy).

(c)*Iniziali nonparametriche basate su polinomi.*

La proposta di Regazzini e Sazonov e' come si e' visto molto generale; in sostanza, essi mostrano che qualunque opinione iniziale su una misura di probabilità aleatoria può essere espressa mediante una distribuzione iniziale del tipo (3), purche' si sia in grado di esprimere la probabilità iniziale dei piccoli ipercubi Δ_ν. Una versione preliminare della proposta di Regazzini e Sazonov può essere quella di approssimare la *distribuzione iniziale* π mediante polinomi di Bernstein.

Una proposta per costruire iniziali nonparametriche, che fa anch'essa uso dei polinomi di Bernstein, e' suggerita da Petrone [9]. La differenza e' che Petrone suggerisce di descrivere non l'iniziale ma il *modello* $\tilde{F}$ come un polinomio di Bernstein (aleatorio). Il lavoro di Petrone si limita al caso di spazio campionario $\mathcal{X} = [0, 1]$, ma sono possibili generalizzazioni, anche per dati multidimensionali (si veda Epifani, 1999). Poiche' assegnare una distribuzione di probabilità iniziale sulla classe $\Delta[0, 1]$ di tutte le funzioni di ripartizione su $[0, 1]$ e' difficile, l'idea e' di descrivere la funzione di ripartizione come un polinomio di Bernstein, e quindi probabilizzare i coefficienti e l'ordine del polinomio, ottenendo un polinomio aleatorio. Si può mostrare che la costruzione e' generale, ovvero fornisce una legge di probabilità che ha come supporto $\Delta[0, 1]$, sfruttando il fatto che ogni funzione di ripartizione su $[0, 1]$ può essere approssimata mediante un polinomio di Bernstein. La scelta dei polinomi di Bernstein rispetto ad altri polinomi e' interessante perche' essi hanno una natura probabilistica; infatti, il polinomio di Bernstein di ordine k che approssima una funzione di ripartizione $F(x)$ e' definito da $B(x; k, F) = E(F(Y/k))$, dove Y ha distribuzione binomiale di parametri k e x. Si può mostrare che se F e' una funzione di ripartizione (con $F(0) = 0$), anche $B(x; k, F)$ e' una funzione di ripartizione (assolutamente continua) su $[0, 1]$; la densità di $B(x; k, F)$ risulta essere una combinazine lineare finita di densità beta di parametri fissati. Ciò collega la proposta con l'idea di descrivere il modello statistico come una combinazione lineare di funzioni base, tecnica utilizzata in letteratura nella stima di densità e nella regressione non parametrica. L'utilizzo dei polinomi di Bernstein aleatori permette di introdurre incertezza sui coefficienti e sul numero di termini della mistura (che corrisponde al parametro di lisciamento) e di fare inferenza su di essi mediante il calcolo della loro distribuzione finale (si veda Petrone, [8]).

4 Determinazione della distribuzione finale e della distribuzione di funzionali.

Nell'impostazione bayesiana, il problema inferenziale e' risolto nel calcolo della distribuzione finale del parametro di interesse, o di una sua sintesi calcolata in base ad una funzione di danno assegnata.

In ambito non parametrico, quindi, una volta assegnata la distribuzione iniziale sul parametro infinito dimensionale, si dovrà calcolare la distribuzione finale. Purtroppo, non sempre tale calcolo e' agevole. Una ragione della popolarità del processo di Dirichlet come iniziale non parametrica nell'inferenza bayesiana e' proprio la semplicità della regola di aggiornamento corrispondente; infatti, e' noto che, se F e' un processo di Dirichlet di parametro α, allora, subordinatamente ad un campione $X_1, \ldots, X_n$ da F, la finale di F e' ancora un processo di Dirichlet, di parametro $\alpha + nF_n$, dove F_n e' la misura di probabilità corrispondente alla funzione di ripartizione empirica di $(X_1, \ldots, X_n)$. Tuttavia, anche per il processo di Dirichlet si presentano difficoltà computazionali, per esempio nel calcolo della distribuzioni di funzionali della F (questo problema sarà trattato nel seguito del paragrafo).

Le difficoltà computazionali possono essere risolte essenzialmente secondo tre vie: fornendo una approssimazione numerica; mediante simulazione stocastica; fornendo delle approssimazioni asintotiche.

Approssimazione numerica. Una approssimazione numerica, con errore di approssimazione assegnato a piacere, può essere ottenuta utilizzando la tecnica di Regazzini e Sazonov [2 e 3]. Tuttavia, tale possibilità sembra per ora più teorica che pratica, in quanto richiede di calcolare una somma su un numero di componenti molto elevato, pari al numero di piccoli ipercubi utilizzati. Tale procedura e' illustrata da Guglielmi e Melilli [5] per dati dicotomici parzialmente scambiabili.

Simulazione stocastica. Un'ampia discussione della possibilità di approssimare le inferenze bayesiane in ambito nonparametrico mediante tecniche di simulazione e' contenuta in un recente lavoro di Walker *et al.* (1998). Le tecniche suggerite sono essenzialmente di tipo Monte Carlo, o Monte Carlo basato su catene di Markov (MCMC). Petrone ([8] e [9]) suggerisce un algoritmo di tipo MCMC per approssimare la finale (e suoi funzionali) corrispondente ad una iniziale di Bernstein-Dirichlet. Tale distribuzione iniziale corrisponde ad esprimere il modello statistico come una mistura di densità beta, con coefficienti e numero di componenti aleatori. Come e' noto, nei modelli mistura la forma analitica della verosimiglianza e' complicata, e questo rende difficile il calcolo della distribuzione finale.

Inoltre, quando il numero di componenti della mistura e' incognito, e' difficile utilizzare gli usuali algoritmi di tipo MCMC, poiche' la dimensione dello spazio parametrico "cambia" al variare del numero di componenti. Tale problema e' usualmente risolto utilizzando algoritmi di tipo *reversible jump* MCMC (Green, 1995); Petrone propone invece una più attenta parametrizzazione, che consente di utilizzare un algoritmo di tipo Monte Carlo con variabili ausiliarie.

Approssimazione asintotica. Le tecniche di simulazione consentono dunque di approssimare la distribuzione finale in molte situazioni. Tuttavia, se l'ampiezza campionaria e' elevata, tali tecniche possono risultare computazionalmente molto onerose. Ci sembra allora importante fornire delle approssimazioni asintotiche. In ambito parametrico, un noto risultato e' il teorema di Bernstein-Von Mises, che stabilisce, sotto opportune condizioni, la normalità asintotica della distribuzione finale del parametro θ. In ambito non parametrico, avendo un parametro infinito-dimensionale, occorrerà stabilire una versione infinito-dimensionale di tale teorema; ovvero, dimostrare la convergenza della legge di probabilità finale ad un processo gaussiano. Nella letteratura bayesiana non parametrica non vi e' stata finora molta attenzione ai problemi asintotici. Alcuni lavori considerano il problema della consistenza della distribuzione finale (si vedano per esempio Diaconis e Freedman, 1986; Ghosal *et al.*, 1997; Shen e Wasserman, 1998). La normalità asintotica in ambito non parametrico e' stata studiata da Lo (1983) per il processo di Dirichlet, e da Cox (1993), Diaconis e Freedman (1997), Conti [1]. Conti [1] propone una analisi bayesiana nonparametrica per un problema di studio delle code. Nel modello di Conti, non vengono formulate ipotesi parametriche sulla distribuzione del tempo di servizio, che si suppone retta da un processo di Dirichlet. In particolare, si e' interessati alla funzione generatrice di probabilità W della distribuzione del tempo d'attesa di equilibrio. Purtroppo, non si riesce a determinare analiticamente la legge di probabilità finale di W; la proposta di Conti e' di utilizzare una sua approssimazione asintotica, giustificata anche dal fatto che, in tale tipo di applicazioni, l'ampiezza campionaria e' solitamente elevata. Viene quindi fornito un risultato di tipo Bernstein-von Mises infinito-dimensionale, che mostra la convergenza della finale di una opportuna versione standardizzata di W ad un processo gaussiano.

4.1 Distribuzione della media di un processo di Dirichlet.

In molte circostanze e' interessante studiare la distribuzione del funzionale $\Gamma_\alpha = \int x\tilde{F}(dx)$, dove $\tilde{F}$ e' un processo di Dirichlet di parametro α.

L'espressione analitica della distribuzione di Γ_α e' stata ricavata da Cifarelli e Regazzini (1990). Precisamente, tali autori ricavano la trasformata generalizzata di Stieltjes di Γ_α e dall'inversione della trasformata ottengono l'espressione della distribuzione. Un modo più diretto di quello seguito da Cifarelli e Regazzini e' stato proposto da Guglielmi (1998*a*) sfruttando la caratterizzazione del processo di Dirichlet data da Sethuraman (1994).

Regazzini [10] suggerisce un diverso procedimento per ottenere la trasformata di Stieltjes, a partire da un risultato di Carlson (1977) che mostra come alcune importanti funzioni speciali possano essere rappresentate come *Dirichlet averages*, cioe' come medie di date funzioni elementari rispetto ad una distribuzione di Dirichlet. Regazzini mostra che la trasformata di Stieltjes di Γ_α può essere rappresentata come una *Dirichlet average* rispetto alla misura di Dirichlet infinito-dimensionale. Il passaggio dal finito-dimensionale all'infinito-dimensionale e' effettuato mediante una tecnica di approssimazione. La tecnica di Carlson ha relazioni con le funzioni di Lauricella, e ciò consente a Regazzini di utilizzare la teoria delle funzioni di Lauricella per ottenere una nuova rappresentazione della trasformata di Stieltjes di Γ_α e di ottenere in modo più semplice e diretto i risultati inerenti alla distribuzione di Γ_α.

Purtroppo, la distribuzione di Γ_α non e' facile da valutare numericamente. Per tale motivo, ne sono state proposte diverse approssimazioni. Muliere e Secchi (1996) hanno suggerito una procedura di bootstrap detta "bootstrap bayesiano proprio". Guglielmi (1998*b*) ha proposto di approssimare la distribuzione di Γ_α con la distribuzione della media di un opportuno processo di Dirichlet con parametro di supporto finito. Due ulteriori proposte, basate su due diverse costruzioni del processo di Dirichlet, sono suggerite da Muliere e Tardella (1998) e Guglielmi [3]. Muliere e Tardella si basano sulla costruzione di Sethuraman (1994); Guglielmi su quella data da Feighin e Tweedie (1989). Sia P_α un processo di Dirichlet di parametro α. La costruzione di Sethuraman può essere scritta come $P_\alpha(\cdot) = \sum_{i=1}^{\infty} p_i \delta_{X_i}(\cdot)$, dove δx indica la misura di probabilità degenere in x, $p_1 = 1 - Y_1$, $p_j = (1 - Y_j)Y_{j-1} \cdots Y_1, j \geq 2$, X_j e Y_j sono v.a. indipendenti; X_j ha distribuzione $\alpha(\cdot)/\alpha(\Re)$ e Y_j ha distribuzione beta di parametri $(\alpha(\Re), 1)$, $\beta(\cdot; \alpha(\Re), 1)$. Muliere e Tardella (1998) definiscono una misura di probabilità aleatoria P_ϵ che e' la somma di n_ϵ termini della serie (assegnando la restante massa di probabilità ad un punto aleatorio), dove n_ϵ e' una v.a. che introduce quindi una regola di arresto. La procedura di approssimazione proposta consiste nel generare campioni Monte Carlo di P_ϵ, per ciascuno dei quali viene valutato il funzionale di P_α di interesse; la distribuzione empirica dei campioni Monte Carlo del funzionale cosi ottenuti approssima la distribuzione di probabilità del funzionale.

La proposta di Guglielmi [3] utilizza invece la costruzione del processo di Dirichlet di Feigin e Tweedie (1989). Sia $\{(X_n, Y_n), n = 1, 2, \ldots\}$ una successione di vettori aleatori indipendenti e identicamente distribuiti, tali che X_n e Y_n sono indipendenti, $X_n \sim \alpha(\cdot)/\alpha(\Re)$ e $Y_n \sim \beta(\alpha(\Re), 1)$. Sia P_0 una misura di probabilità su $\Re$ e $P_n = (1 - Y_n)\delta_{X_n} + Y_n P_{n-1}$ per $n \geq 1$; allora, $\{P_n, n \geq 0\}$ e' una catena di Markov la cui distribuzione invariante e' un processo di Dirichlet di parametro α. Inoltre, la successione delle medie

$$\Gamma_n = \int x dP_n(dx) = (1 - Y_n)X_n + Y_n \Gamma_{n-1}$$

costituisce anch'essa una catena markoviana che sotto condizioni di regolarità ha come distribuzione invariante la legge di probabilità π_α di $\Gamma_\alpha = \int x dP_\alpha(x)$. Guglielmi studia la convergenza della catena markoviana Γ_n, fornendo un confine quantitativo per la distanza variazionale fra la legge di probabilità di Γ_n e la distribuzione stazionaria π_α, dato il valore iniziale Γ_0. Propone quindi di approssimare la legge di probabilità π_α della media Γ_α con la distribuzione empirica di un campione estratto da Γ_n.

5 Strutture di dipendenza più complesse: scambiabilità parziale

La scambiabilità e' lo schema di base nell'inferenza bayesiana, e corrisponde alla situazione di esperimenti omogenei. In molte applicazioni, tuttavia, può essere più realistico pensare a strutture di dipendenza più complesse fra le osservazioni. Nel caso di dati eterogenei, per esempio, si può ritenere appropriata un'ipotesi di scambiabilità parziale. In questo paragrafo discuteremo alcune proposte relative all'analisi bayesiana non parametrica di dati parzialmente scambiabili. La nozione di scambiabilità parziale e' stata introdotta da de Finetti (1938) e ripresa, da una diversa angolazione, da Diaconis e Freedman (1980). Si può dimostrare che la definizione di scambiabilità parziale di Diaconis e Freedman (unita ad una condizione di ricorrenza) e' equivalente a quella di de Finetti (Regazzini *et al.*, 1999).

Analogamente a quanto fatto nel caso di successioni scambiabili, daremo ora i teoremi di rappresentazione per successioni parzialmente scambiabili, discutendo quindi il problema dell'assegnazione della legge di probabilità iniziale, secondo l'approccio previsivo diretto e secondo l'approccio indiretto. Un problema nuovo che si pone nel caso di parziale scambiabilità e' come vedremo lo studio dei legami di dipendenza fra i gruppi.

Parziale scambiabilità alla de Finetti. Sia $X = \{X_{n,m},\ n = 1, 2, \ldots; m = 1, 2, \ldots, k\}$ una successione di variabili aleatorie a valori in $\mathcal{X} \subseteq \Re^d$ (si potrebbe tuttavia considerare il caso più generale di uno spazio polacco). La successione X si dice parzialmente scambiabile se le successioni $\{X_{n,1}\}, \ldots, \{X_{n,k}\}$ sono ognuna una successione scambiabile. In base al teorema di rappresentazione (de Finetti, 1937; Regazzini, 1991), X e' parzialmente scambiabile se e solo se la sua legge di probabilità può essere espressa come

$$P(X_{1,1} \le x_{1,1}, \ldots, X_{n_1,1} \le x_{n_1,1}, \ldots, X_{1,k} \le x_{1,k}, \ldots, X_{n_k,k} \le x_{n_k,k}) =$$

$$\int_{\mathcal{F}^k} \prod_{i=1}^{n_1} F_1(x_{i,1}) \cdots \prod_{i=1}^{n_k} F_k(x_{i,k})\, \pi(dF_1, \ldots, dF_k) \tag{4}$$

per ogni $(x_{1,1}, \ldots, x_{n_k,k})$ e $n_1, \ldots, n_k \ge 1$. Inoltre, la successione dei vettori delle funzioni di ripartizione empiriche $(F_{n,1}, \ldots, F_{n,k})$ (dove $F_{n,j}(x) = \frac{\sum_{i=1}^n \delta_{(-\infty,x]}(x_{i,j})}{n}$), converge debolmente a $(\tilde{F}_1, \ldots, \tilde{F}_k)$. La legge di probabilità π e' la legge di probabilità di $(\tilde{F}_1, \ldots, \tilde{F}_k)$, ed e' detta misura di de Finetti della successione X, o distribuzione iniziale. La realizzazione di $\tilde{F}_j$ e' detta modello statistico del gruppo j-mo, $j = 1, 2, \ldots, k$.

Il problema che si pone e' come assegnare la legge di probabilità iniziale π, tenendo conto che essa stabilisce il legame di dipendenza fra i gruppi. In ambito non parametrico, assegnare π significa assegnare una legge di probabilità al vettore di funzioni di ripartizione aleatorie $(\tilde{F}_1, \ldots, \tilde{F}_k)$. Cifarelli e Regazzini (1978) hanno proposto una iniziale detta mistura di prodotti di processi di Dirichlet. Essa consiste essenzialmente nell'assegnare la distribuzione di $(\tilde{F}_1, \ldots, \tilde{F}_k)$ in modo gerarchico, assumendo che $\tilde{F}_1, \ldots, \tilde{F}_k$ siano condizionatamente indipendenti dato un vettore di parametri $\theta = (\theta_1, \ldots, \theta_k)$, con $(\tilde{F}_j \mid \theta) \sim \mathcal{D}(\alpha_{\theta_j})$, cioe' un processo di Dirichlet di parametro α_{θ_j}; θ e' un parametro aleatorio con funzione di ripartizione H cos'ì che

$$(\tilde{F}_1, \ldots, \tilde{F}_k) \sim \int \prod_{j=1}^{k} \mathcal{D}(\alpha_{\theta_j}) H(d\theta) .$$

Guglielmi e Melilli [5] hanno mostrato che ogni distribuzione iniziale per successioni parzialmente scambiabili può essere approssimata da una mistura di prodotti di misure di Dirichlet. La costruzione di tale mistura si basa sui risultati di Regazzini e Sazonov [12] e [13]. Precisamente, Guglielmi e Melilli costruiscono, data una qualunque iniziale π ed $\epsilon > 0$, una mistura di prodotti di misure di Dirichlet che approssima π (nella metrica di Prohorov), con errore minore di ϵ. Consideriamo in particolare il caso di due successioni parzialmente scambiabili $\{X_n, Y_n, n \geq 1\}$ a valori $\{0;1\}$; il teorema di rappresentazione consente di scrivere:

$$P(X_1 = x_1, \ldots, X_{n_1} = x_{n_1}, Y_1 = y_1, \ldots, Y_{n_2} = y_{n_2}) =$$

$$= \int_{[0,1]^2} \theta_1^{\sum_{i=1}^{n_1} x_i} (1-\theta_1)^{n_1 - \sum_{i=1}^{n_1} x_i} \theta_2^{\sum_{i=1}^{n_2} y_i} (1-\theta_2)^{n_2 - \sum_{i=1}^{n_2} y_i} \pi(d\theta_1, d\theta_2) . \tag{5}$$

In questo caso, la distribuzione iniziale approssimante proposta da Guglielmi e Melilli coincide con il polinomio di Bernstein sul quadrato unitario $[0,1]^2$ che approssima π. Questa costruzione si può estendere al caso generale di spazio campionario polacco.

Parziale scambiabilità alla Diaconis e Freedman. Sia $X = \{X_1, X_2, \ldots\}$ una successione di numeri aleatori a valori in un insieme numerabile I che chiameremo spazio degli stati. Sia $\sigma = (x_0, x_1, \ldots, x_n)$ una stringa finita di elementi di I e $t_{i,j}(\sigma)$ il numero di transizioni da i a j in σ per ogni $i, j \in I$.

Introduciamo la seguente relazione di equivalenza. Se $\sigma = (x_0, \ldots, x_n)$ e $\phi = (y_0, \ldots, y_n)$ sono due stringhe finite, diciamo che σ e' equivalente a ϕ se σ e ϕ hanno lo stesso elemento iniziale e $t_{i,j}(\sigma) = t_{i,j}(\phi)$, per ogni $i, j \in I$. Ad esempio, le stringhe $(0,1,2,0,1,0)$ e $(0,1,0,1,2,0)$ sono equivalenti.

La successione X si dice parzialmente scambiabile alla Diaconis e Freedman se vale

$$P(X_0 = x_0, \ldots, X_n = x_n) = P(X_0 = y_0, \ldots, X_n = y_n)$$

per ogni coppia di stringhe equivalenti σ e ϕ. Diaconis e Freedman (1980) forniscono il seguente teorema di rappresentazione. Supponiamo che la successione X sia ricorrente, ossia che $P(X_n = x_0 \text{ per infiniti } n) = 1$. Allora, la successione ricorrente X e' parzialmente scambiabile alla Diaconis e Freedman se e solo se e' una mistura di catene di Markov; cioe', esiste una misura π sull'insieme $\mathcal{P}$ delle matrici stocastiche su $I \times I$ tale che per ogni $n \geq 1$ e per ogni $(x_0, x_1, \ldots, x_n) \in I^{n+1}$ vale

$$P(X_0 = x_0, \ldots, X_n = x_n) = \int_{\mathcal{P}} \prod_{j=0}^{n-1} s(x_j, x_{j+1}) \pi(ds) . \tag{6}$$

In altre parole, condizionatamente alla matrice di transizione s, X e' una catena di Markov con matrice di transizione s. La distribuzione iniziale π di s può essere caratterizzata in termini puramente previsivi; per esempio, Muliere, Secchi and Walker [6] hanno proposto uno schema d'urna detto *processo d'urna rinforzato* che genera misture di catene di Markov tali che la legge di s e' il prodotto di processi di Dirichlet. E' possibile pensare ad una estensione di tale risultato rendendo aleatoria la composizione dell'urna, che porti a caratterizzare l'iniziale come mistura di prodotti di processi di Dirichlet. Ci si può quindi aspettare, sulla base dei risultati di Guglielmi e Melilli [5] e dell'equivalenza fra la scambiabilità nel senso di Diaconis e Freedman e di de Finetti, che ogni legge di probabilità parzialmente scambiabile per successioni a valori discreti possa essere approssimata mediante un processo d'urna rinforzato. Un caso particolare di processi d'urna rinforzati sono i Polya trees (Mauldin *et al.*, 1992).

Illustriamo il processo d'urna rinforzato in un semplice esempio. Sia $I = \{0, 1, \ldots\}$ e supponiamo di avere una successione di urne contenenti palline bianche e palline nere. Assumiamo che l'urna zero $U(0)$ contenga solo palline nere e l'urna $x-$esima $U(x)$, per $x \geq 1$, contenga $n_x(b) > 0$ palline bianche e $n_x(n) > 0$ palline nere. Sia $X_0 = 0$; il processo X_n si muove secondo la seguente *legge di moto*: se $X_{n-1} = x$, estraiamo una pallina dall'urna $U(x)$ e la riponiamo nell'urna insieme ad una pallina dello stesso colore; quindi, $X_n = x + 1$ se la pallina estratta e' nera, $X_n = 0$ se la pallina e' bianca. Muliere e Secchi [7] mostrano che:

* il processo $\{X_n, n \geq 0\}$ e' ricorrente se e solo se $lim_{n\to\infty} \prod_{i=0}^{n} \frac{n_x(n)}{n_x(n)+n_x(b)} = 0$;

* $\{X_n, n \geq 0\}$ e' parzialmente scambiabile; quindi vale il teorema di rappresentazione di Diaconis e Freedman (1980);

* se $\{X_n, n \leq 0\}$ e' ricorrente, allora la matrice di transizione $\tilde{s}$ nella rappresentazione (6) e' tale che: $\tilde{s}(i, i+1) = 1 - \tilde{s}(i, 0)$ con probabilità uno, per ogni $i \in I$; inoltre, $\tilde{s}$ ha elementi $\tilde{s}(i, i+1)$ indipendenti, con $\tilde{s}(i, i+1)$ avente distribuzione beta di parametri $(n_i(b), n_i(n))$.

Inoltre, Muliere, Secchi e Walker mostrano che, per dati processi d'urna rinforzati, si possono definire delle funzioni degli osservabili che risultano essere scambiabili; la misura di de Finetti di tali successioni scambiabili e' un processo beta-Stacy (Walker e Muliere, 1997*a*) o, più in generale, un processo neutrale a destra (Doksum, 1974). Con riferimento all'esempio precedente, definiamo i tempi d'arresto $\tau_1 = \inf\{n > 0 : X_n = 0\}$ e per ogni $n \geq 2$, $\tau_n = \inf\{n > \tau_{n-1} : X_n = 0\}$. Per ogni $n \geq 1$, definiamo $T_n = X_{\tau_n - 1}$. Si dimostra che, se $\{X_n\}$ e' ricorrente, la successione $\{T_n\}$ e' scambiabile, e la misura di de Finetti di $\{T_n\}$ e' un processo beta-Stacy.

I processi d'urna rinforzati hanno applicazioni immediate nell'analisi della sopravvivenza, e possono essere generalizzati per trattare problemi multistato.

Studio della dipendenza fra i gruppi. Come accennato, nel caso di dati parzialmente scambiabili la distribuzione iniziale stabilisce il legame di dipendenza fra i gruppi. Ad esempio, nel caso di successioni di eventi rappresentato dalla (5), se θ_1 e θ_2 sono indipendenti, abbiamo due gruppi di osservazioni scambiabili al loro interno e indipendenti fra loro; all'estremo opposto, se la distribuzione di (θ_1, θ_2) e' tale che $\theta_1 = \theta_2$ con probabilità uno, si ritorna alla scambiabilità totale.

Nel caso più generale (4) (in cui assumiamo $k = 2$), si può allora pensare di misurare la "distanza" fra la situazione di dipendenza fra le osservazioni indotta da una iniziale π e quella di scambiabilità totale, mediante un indice della differenza fra i modelli F_1 e F_2. Tale idea e' sviluppata da Guglielmi e Melilli [4], che propongono di utilizzare un indice del tipo $I(\pi) = \int_{\mathcal{F}} \sigma(F_1, F_2) d\pi(F_1, F_2)$, dove $\sigma(F_1, F_2)$ e' un indice di differenza fra le distribuzioni F_1 e F_2, che e' nullo se e solo se $F_1 = F_2$. Purtroppo, tale indice e' difficile da calcolare, a meno che l'iniziale sia di tipo parametrico. Guglielmi e Melilli propongono pertanto un altro indice, basato sul confronto fra le distribuzioni finito-dimensionali della successione e quelle di una successione totalmente scambiabile. L'indice proposto risulta anch'esso difficile da calcolare, poiche' coinvolge distribuzioni di dimensione elevata. Una proposta alternativa consiste allora nel confrontare, anziche' le leggi finito-dimensionali, le distribuzioni di funzioni significative delle osservazioni. La proposta e' di considerare a tal fine le distribuzioni di riassunti esaustivi a fini previsivi.

6 Note e commenti

Nel 1972 D.V. Lindley scriveva : *"Nonparametric Statistics. This is a subject about which the Bayesian method is embarassingly silent"*. Da allora la statistica bayesiana non parametrica ha fatto considerevoli passi in avanti sia da un punto di vista teorico, sia da un punto di vista computazionale (per una recente rassegna si veda Walker *et al.*, 1999). A nostra conoscenza, la prima precisa formulazione bayesiana di un problema non parametrico fu proposta da Bruno de Finetti (1935) in un lavoro riguardante il problema della perequazione, ovvero il problema della sostituzione di una "distribuzione teorica" a dati empirici. Egli propose due possibili modi per affrontare la questione posta, il secondo dei quali coincide appunto con ciò che oggi e' noto come impostazione bayesiana non parametrica. De Finetti affrontò il problema all'interno dello schema di scambiabilità , ed e' questo lo schema di base che abbiamo adottato in questa relazione per presentare i contributi del gruppo.

All'epoca in cui de Finetti proponeva il procedimento, mancavano completamente proposte di processi che potessero essere utilizzati come distribuzioni iniziali. Le prime distribuzioni iniziali si devono a Freedman (1963) e Fabius (1964), che introdussero i processi *tailfree*. E' di Rolph (1968) il primo tentativo diretto di costruire misure di probabilità aleatorie utilizzando i momenti; furono però i contributi di Ferguson (1973, 1974), sulla base dei lavori di Fabius (1964), Dubins e Freedman (1965) e Freedman (1965), che diedero un impulso decisivo alla applicazione delle metologie bayesiane per la soluzione di svariati problemi di inferenza non parametrica. Ferguson introdusse il processo di Dirichlet come misura di probabilità aleatoria su spazi misurabili astratti.

Dopo Ferguson la letteratura bayesiana non parametrica e' cresciuta rapidamente e, come sempre avviene nello sviluppo scientifico, l'esigenza di risolvere nuovi problemi ha portato all'introduzione di nuovi processi. Antoniak (1974) ha considerato una importante generalizzazione del processo di Dirichlet: la mistura di processi di Dirichlet. Per affrontare problemi in condizioni di scambiabilità parziale, Cifarelli e Regazzini (1978) hanno introdotto le misture di prodotti di processi di Dirichlet; tali processi sono stati utilizzati in numerose applicazioni: nel modello lineare (Cifarelli, Muliere e Scarsini, 1981), nell'analisi della varianza (Cifarelli, 1979; Muliere e Scarsini, 1983), nell'analisi discriminante (Consonni, 1981), nella determinazione della dose massima di un farmaco (Muliere e Petrone, 1993), nella cluster analysis bayesiana (Petrone e Raftery, 1997).

Una importante e ricca classe di priors e' data dai processi "neutrali a destra" (Doksum (1974). Molti processi ben noti appartengono a questa classe: il processo omogeneo semplice (Ferguson e Phadia, 1979), il processo di Dirichlet, il processo gamma (Ferguson, 1974). Recentemente e' stato introdotto il processo beta-Stacy (Walker e Muliere, 1997*a*), che e' un processo neutrale a destra che generalizza i precedenti. Tali processi sono particolarmente utili nell'analisi della sopravvivenza, dove in generale ci si trova di fronte a dati censurati.

I processi neutrali a destra sono caratterizzati mediante i processi di Levy ad incrementi indipendenti. Per alcune applicazioni si veda Kalbfleisch (1978), Laud et al. (1998). I processi ad incrementi indipendenti possono essere utilizzati anche per modellare funzioni diverse dalla funzione di ripartizione. Hjort (1990) ha introdotto il processo Beta per modellare la funzione di *hazard* cumulativa. Dykstra e Laud (1981) hanno proposto di modellare il *monotone hazard rate* non parametricamente utilizzando il processo gamma esteso e Arjas a Gasbarra (1994) hanno sviluppato altri processi per modellare *hazard rate piecewise*.

Un posto di rilievo hanno infine le misure di probabilità aleatorie costruite a partire da alcuni schemi d'urna (del tipo Polya). Contributi recenti in questa direzione sono le proposte di Mauldin et al. (1992) e Lavine (1992,1994), che hanno introdotto i Polya trees. Sulla stessa linea di ricerca si colloca il lavoro di Muliere, Secchi e Walker[6], che hanno proposto i processi d'urna rinforzati quale generalizzazione dei Polya trees (si veda il paragrafo 5). Tali processi hanno notevoli possibilità applicative nell'analisi della sopravvivenza. Una diversa generalizzazione dei Polya trees e' dovuta a Monticino (1997) che ha introdotto i cosidetti *trees exchangeable processes.*

Una distribuzione iniziale nonparametrica basata su polinomi, che seleziona funzioni di ripartizione continue, e' stata proposta da Petrone ([9]; si veda il paragrafo 3).

7 Riferimenti bibliografici

Contributi dei componenti del gruppo strategico CNR, discussi in questa relazione

[1] Conti P.L. (1998) Large sample Bayesian analysis for Geo/G/1 discrete-time queueing models. (Mimeo)

[2] Fortini, S., Ladelli, L. e Regazzini, E. (1998). Exchangeability, predictive distributions and parametric models. *Technical Report* 98.5, IAMI-CNR, Milano.

[3] Guglielmi, A. (1998), Rate of convergence to the law of the mean of a Dirichlet process. *Technical Report* 98.18, IAMI-CNR, Milano.

[4] Guglielmi A, e Melilli,E. (1998), Measuring exchangeability in a partially exchangeable sequence. *Technical Report* 98.10, IAMI-CNR, Milano.

[5] Guglielmi,A e Melilli, E. (1998), Approximating de Finetti's measures for partially exchangeable sequences. *Technical Report* 98.17, IAMI-CNR, Milano.

[6] Muliere, P., Secchi, P. e Walker, S. (1998). Urn schemes and reinforced random walks for Bayesian nonparametrics, *Quaderni di Dipartimento* # 71(2-98), Dip. Econ. Pol. Met. Quant., Università di Pavia.

[7] Muliere, P. e Secchi, P. (1998). Uno schema d'urna per il processo beta-Stacy. *Atti XXXIX Riunione Scientifica della SIS*, in pubblicazione.

[8] Petrone, S. (1998). Bayesian density estimation using Bernstein polynomials, *Canad. J. Statist.*, in pubblicazione.

[9] Petrone, S. (1998). Random Bernstein polynomials, *Scandinav. J. Statist.*, in pubblicazione.

[10] Regazzini, E. (1997). An example of the interplay between statistics and special functions. *Technical Report* 97.22, IAMI-CNR, Milano.

[11] Regazzini, E. (1998). Old and recent results on the relationship between predictive inference and statistical modelling either in nonparametric or parametric form. In pubblicazione in *Bayesian Statistics 6*, Bernardo et. al. eds., Oxford University Press.

[12] Regazzini, E. e Sazonov V.V. (1997). Approximation of laws of random probabilities by mixtures of Dirichlet distributions with applications to nonparametric Bayesian inference. Part 1. *Technical Report* 97.7 (I), IAMI-CNR, Milano.

[13] Regazzini, E. e Sazonov V.V. (1997). Approximation of laws of random probabilities by mixtures of Dirichlet distributions with applications to nonparametric Bayesian inference. Part 2. *Technical Report* 97.7 (II), IAMI-CNR, Milano.

Bibliografia

Altomare, F. e Campiti, M. (1994). *Korovkin type approximation theory and its apupliction.* W. de Gruyter, Berlin.

Andreev, A. e Arjas, E. (1996). A note on histogram aproximation in Bayesian density estimation. In *Bayesian Statistics 5* (J.M.Bernardo, J.O. Berger, A.P. David and A.F.M. Smith, Eds) Oxford:Univeristy Press, Oxford, 487-490.

Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.*, **2**, 1152-1174.

Arjas, E. e Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statistica Sinica*, **4**, 505-524.

Blackwell, D. e MacQueen, J.B. (1973). Ferguson distributions via Polya-urn schemes. *Ann. Statist.* **1**, 353-355.

Campanino, D.M. e Spizzichino, F. (1981). Prediction sufficiency and representation of infinite sequences of exchangeable random variables. *Quaderni dell' Istituto Matematico "G. Castelnuovo"*, Università di Roma.

Carlson, B.C. (1977). *Special functions of applied mathematics.* Academic Press, New York.

Cifarelli, D.M. (1979). Impostazione bayesiana di un problema di analisi della varianza con approccio non parametrico. *Quaderni Istituto di Matematica Finanziaria* , Università di Torino.

Cifarelli, D.M. e Regazzini,E.(1978). Problemi statistici non parametrici in condizioni di scambiabilità parziale impiego di medie associative.*Quaderni Istituto di Matematica Finanziaria* Università di Torino.

Cifarelli, D.M. e Regazzini, E. (1980). Sul ruolo dei riassunti esaustivi ai fini della previsione in un contesto bayesiano (parte I). *Rivista di Matematica per le Scienze Economiche e Sociali*, **3**, 109-125.

Cifarelli, D.M. e Regazzini, E. (1981). Sul ruolo dei riassunti esaustivi ai fini della previsione in un contesto bayesiano (parte II). *Rivista di Matematica per le Scienze Economiche e Sociali*, **3**, 3-11.

Cifarelli, D.M., Muliere, P. e Scarsini, M. (1981). Il modello lineare nell'approccio Bayesiano non parametrico. *Quaderni dell'Istituto Matematico "G. Castelnuovo"* Università degli Studi, Roma.

Cifarelli, D.M. e Regazzini, E. (1990). Distribution functions of means of a Dirichlet process. *Ann. Statist.*.**18**, 429-442 (correction in *Ann. Statist.*(1994),**22**, 1633-1634.

Consonni, G. (1981), Impostazione Bayesiana di un problema di analisi discriminatoria nell'ambito di un modello non parametrico. *Rivista di matematica per le Scienze Economiche e Sociali* 4, 89-102.

Cox, D. (1993). An analysis of Bayeisna inference for nonparametric regression. *Ann. Statist.*, **21**, 903-923.

Dalal, S.R. (1978). A note on the adeguacy of mixtures of Dirichlet processes. *Sankhya*, **40**, 185-191.

Dalal, S.R. e Hall,G.J. (1980). On approximating parametric Bayes models by nonparametric Bayes models. *Ann. Statist.***8**, 664-672.

Dalal, S.R. e Hall, W.J. (1983). Approximating priors by mixtures of natural conjugate priors. *J. Roy. Statist. Soc. Ser. B*, **45**, 278-286.

de Finetti, B. (1935), Il problema della perequazione. *Atti della società Italiana per il Progresso delle Scienze, (XII Riunione)*, Napoli.

de Finetti, B. (1937). La prevision: ses lois logiques, ses sources subjectives. *Ann. Instit. H. Poincare* 7, 1-68.

de Finetti, B. (1938). Sur la condition d' equivalence partielle. VI Colloque Geneve. *Act. Sci. Ind.*, 739, 5–18.

Diaconis, P. e Ylvisaker, D. (1985). Quantifying prior opinion. *Bayesian Statistics 2*, J.M. Bernardo, M.H. deGroot, D.V. Lindley, A.F.M. Smith Eds., Elsevier Science Publisher B.V. (north Holland), 133-156.

Diaconis, P. e Freedman, D. (1980). de Finetti's theorem for Markov chains. *Ann. Probab.*, **8**, 115–130.

Diaconis, P. e Freedman, D. (1986). On the consistency of bayes estimates. *Ann. Statist.*, **14**, 1-26.

Diaconis, P. e Freedman, D. (1997). On the Bernstein-von Mises theorem with infinite dimensional parameter.. *Technical Report 492*, Dept. of Statistics, University of California at Berkeley.

Doksum, K.A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.*,**2**, 183-201.

Dubins, L. e Freedman, D. (1965). Random distribution functions. *Bull. Amer. Math. Soc.*, **69**, 548-551.

Dykstra, R.L. e Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356-367.

Epifani, I. *Alcuni risultati sulle leggi di probabilità dei polinomi aleatori di Bernstein e loro applicazioni alla statistica bayesiana.* Tesi di Dottorato in Statistica Metodologica (XI ciclo), Università di Trento.

Fabius, J. (1964). Asymtotic Behavior of Bayes estimate. *Ann. Math. Statist.* **35**, 846-856.

Feigin e Tweedie (1989). Linear functionals and markov chains associated with Dirichlet processes. *Math. Proc. Phil. Soc.*, **105**, 579-585.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.*, **2**, 615-629.

Ferguson, T.S. e Phadia, E.G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.*, **7**, 163-186.

Freedman, D.A. (1963). On the asymptotic behavior of Bayes' estimate in the discrete case. *Ann. Math. Statist.* **34**, 1386-1403.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**, 711–732.

Ghosal, S., Gosh, J.K., Ramamoorthy, R.V. (1997). Consistency issues in Bayesian nonparametrics. In preparazione.

Guglielmi, A. (1998*a*). A simple procedure calculating the generalized Stieltjes transform of the mean of the Dirichlet process. *Statist. Prob. Letters.* **38**, 299-303.

Guglielmi, A. (1998*b*. Numerical analysis for the distribution function of the mean of a Dirichlet process. *Quaderno IAMI 98.01*, CNR-IAMI, Milano.

Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.*, **18**, 1259-1294.

Kalbfleish, J.D. (1978). Nonparametric Bayesian analysis of survival time data. *J. Roy. Statist. Soc. Ser. B*, **40**, 214-221.

Krasker, W.S. (1984). A note on selecting parametric models in Bayesian inference. *Ann. Statist.*, **12**, 751-757.

Laud, P.W., Damien, P. e Smith, A.F.M. (1998). Bayesian nonparametric and covariate analysis of failure time data. In *Practical nonparametric and semiparametric Bayesian statistics*, D. Dey, P. Mueller e D. Sinha (Eds.), Springer, New York, 213-225.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *ann. Statist.*, **20**, 1203-1221.

Lavine, M. (1994). More aspects of Plya trees for statistical modelling. *Ann. Statist.*, **22**, 1164-1176.

Lindley, D.V.(1972) *Bayesian Statistics: A Review*, Society for Industrial and Applied Mathematics, Philadelphia.

Lo, A.Y.(1983). Weak convergence for Dirichlet process. *Sankhya*, **45**, 105-111.

Lo, A.Y. (1991). A characterization of the Dirichlet process. *Statist. Probab. Lett.*, **12**, 185–187.

Monticino, (1997). Constructing prior distributions with trees of exchangeable processes, *Journal of Statistical Planning and Inference* , in pubblicazione.

Mauldin, R.D., Sudderth, W.D. e Williams, S.C. (1992). Polya trees and random distributions. *Ann. Statist.*, **20**, 1203-1221.

Muliere,P and Scarsini M. (1983). Impostazione Bayesiana di un problema di analisi della varianza a due criteri. *Giornale degli Economisti e Annali di Economia*, 519- 526.

Muliere, P. e Petrone, S. (1993). A Bayesian predictive approach to sequential search for an optimal dose: parametric and nonparametric models. *J. Ital. Statist. Soc.*, **3**, 349-364.

Muliere, P. e Secchi, P. (1996). Bayesian nonparametric predictive inference and bootstrap techniques. *Ann. Inst. Statist. Math.*, **48**, 663-673.

Muliere, P. e Walker, S.G. (1997a). A Bayesian nonparametric approach to survival analysis using Polya trees. *Scand. J. Statist.*, **24**, 331-340.

Muliere,P e Tardella, L. (1998). Approximating distributions of random functionals of Ferguson-Dirichlet priors. *The Canadian Journal of Statistics*, **26**, 283-297.

Petrone, S. e Raftery, A.E. (1997). A note on the Dirichlet process prior in Bayesian nonparametric inference with partial exchangeability. *Statist. Prob. Letters*, **36**, 69-83.

Regazzini, E. (1978). Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità. *Giornale dell'Istituto Italiano degli Attuari*, **41**, 77-89.

Regazzini, E. (1991). Coherence, exchangeability and statistical models (de Finetti's stance "revisited"). *Quaderno IAMI 91-11*, IAMI-CNR, Milano.

Regazzini, E., Fortini, S., Ladelli, L. e Petris, G. (1999). On mixtures of distributions of Markov chains. (mimeo).

Rolph, J.E.(1968), Bayesian Estimation of Mixing distributions, *Annal. Math. Statist.*, **39**, 1289- 1302.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**, 639-650.

Shen, X. e Wasserman, L. (1998). Rates of convergence of posterior distributions. *Technical report No. 678*, Carnegie Mellon.

Susarla, V. e Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete data. *J. Amer. Statist. Assoc.*, **71** , 897-902.

Walker, S. e Muliere, P. (1997*a*). A characterization of Pólya tree distributions. *Statist. Probab. Lett.*, **31**, 163–168.

Walker, S.G. e Muliere, P. (1997*b*). Beta-Stacy Processes and a generalisation of the Polya-urn Scheme. *Ann. Statist.*, **25**, 1762- 1780.

Walker, S., Damien, P., Laud, P.W. e Smith, A.F.M. (1999). Bayesian nonparametric inference for random distributions and related functions. (with discussion) *J. Roy. Statist. Soc. Ser. B*, **61**, 485-527.