



UNIVERSITÀ DI PISA

Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea Specialistica in Matematica

Tesi di Laurea

PROCESSI DI DIRICHLET ED APPLICAZIONI

Relatore:
Prof. FRANCO FLANDOLI

Candidato:
LUIGI AMEDEO BIANCHI

Controrelatore:
Prof. GIORGIO LETTA

ANNO ACCADEMICO 2008-2009

La teoria della probabilità è, in fondo, semplice buon senso tradotto in calcolo; ci fa valutare con esattezza ciò che una mente ragionevole sente per una sorta di istinto... È degno di nota che questa scienza, nata a servizio dei giochi d'azzardo, sia diventata il più importante oggetto della conoscenza umana... Le più importanti questioni della vita sono, per la maggior parte, solo dei problemi di probabilità.

P.S. de Laplace

Introduzione

In questa tesi vogliamo dare una presentazione del processo di Dirichlet, fornendone una costruzione rigorosa e sottolineandone le proprietà principali. Sfrutteremo poi l'oggetto così ottenuto per introdurre una particolare misura sullo spazio delle probabilità su una varietà riemanniana compatta: la misura entropica.

Introduciamo, come prima cosa, alcune definizioni e risultati di base, tra cui le distribuzioni di Dirichlet, il teorema di Kolmogorov e le misure aleatorie. Affrontiamo poi la costruzione classica del processo di Dirichlet, mediante il teorema di Kolmogorov, dopo aver presentato la tecnica nel caso del ben noto processo di Wiener, o moto browniano. Un processo di Dirichlet μ sullo spazio \mathcal{X} avente per parametro una misura finita α è tale che se $X \sim \mu(\alpha)$, allora per ogni partizione $\{B_i\}_{i=1}^k$ abbiamo

$$(X(B_1), \dots, X(B_k)) \sim \text{Dir}(\alpha(B_1), \dots, \alpha(B_k)).$$

Entrano allora in gioco le proprietà delle distribuzioni di Dirichlet presentate in precedenza.

Proseguiamo con costruzioni alternative che ci consentono una più immediata comprensione e ci permettono di mettere in evidenza alcune proprietà del processo. In particolare da una di esse segue direttamente il fatto che un'estrazione dal processo di Dirichlet sia quasi certamente una distribuzione discreta.

Passiamo poi alla costruzione di una misura, detta misura entropica, sullo spazio $\mathcal{P}(M)$ delle misure di probabilità su una varietà riemanniana compatta M . Tale misura è definita, formalmente, dalla relazione

$$d\mathbb{P}^\beta(\nu) = \frac{1}{Z} e^{-\beta \cdot \text{Ent}(\nu)} d\mathbb{P}^0(\nu),$$

in cui \mathbb{P}^0 è una (inesistente) misura uniforme su $\mathcal{P}(M)$. L'approccio euristico per la costruzione muove i passi dall'analogia relazione formale per la misura di Wiener,

$$\mathbb{P}^W(dg) = \frac{1}{Z} e^{-H(g)} \mathbb{P}(dg),$$

proposta da Feynman, in cui H è il funzionale dell'energia. Diamo infine una definizione rigorosa della misura entropica come misura immagine del processo di Dirichlet, rispetto ad un'opportuna mappa di coniugio \mathfrak{C} .

Presentiamo poi una proposta di modello per la previsione spaziale di eruzioni in una zona vulcanicamente attiva. Questo pone le sue basi sulla stretta parentela che intercorre tra le distribuzioni beta e il processo di Dirichlet per restituire, per ogni partizione della zona, una misura della probabilità che la successiva eruzione abbia luogo nella zona medesima. Tale modello prende come distribuzione a priori una misura (finita, ma non necessariamente di probabilità) del rischio di eruzione fornita dagli esperti, per poi aggiornarla con i dati osservativi. Di quest'idea viene proposta una realizzazione in un caso particolare, quello dei Campi Flegrei.

Il nostro lavoro si basa, per la costruzione presentata al capitolo 3, sull'articolo di Ferguson [Fer73], per le costruzioni del capitolo 4 agli articoli di Blackwell e MacQueen [BM73] e di Sethuraman [Set94]. L'introduzione e la definizione della misura entropica sono contenute negli articoli di Sturm e von Renesse [vRS08] e Sturm [Stu09].

Indice

1	Preliminari	1
1.1	Distribuzione di Dirichlet	1
1.2	Teorema di Kolmogorov	4
1.3	Misure aleatorie	5
2	Il processo di Wiener	9
2.1	Costruzione	9
2.2	Misura di Wiener	14
3	Il processo di Dirichlet	17
4	Altri approcci	27
4.1	Urne di Pólya	27
4.2	Stick-breaking	31
4.2.1	μ è la misura di Dirichlet	32
4.3	Ristorante Cinese	34
5	Misura entropica	37
5.1	Gli spazi	38
5.2	Costruzione	41
5.3	Ulteriori considerazioni	45
6	Campi Flegrei	47
6.1	Modello a due zone	47
6.2	Modello a $d + 1$ zone (d -dimensionale)	49
6.3	Modello continuo	50
6.4	Il modello in azione	51
6.5	Conclusioni	52

Capitolo 1

Preliminari

In questo primo capitolo introduciamo alcuni degli strumenti basilari che ci occorreranno in seguito per la trattazione del processo di Dirichlet. In particolare vedremo le distribuzioni di Dirichlet e le loro importanti proprietà, un'introduzione alle misure aleatorie e il teorema di Kolmogorov.

1.1 Distribuzione di Dirichlet

In questa prima sezione definiamo la distribuzione di Dirichlet e ne richiamiamo alcune importanti proprietà, che ci serviranno in seguito per caratterizzare il processo di Dirichlet. Cominciamo allora dalla definizione di distribuzione beta, di cui la Dirichlet è, come vedremo, una generalizzazione.

Definizione 1.1. La *distribuzione beta* è la distribuzione continua definita dalla seguente densità di probabilità:

$$f(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1} \quad (1.1)$$

per $x \in [0, 1]$, con $f(x) = 0$ altrove. I parametri α_1 e α_2 sono reali positivi. In seguito denoteremo la distribuzione di probabilità avente la 1.1 per densità come $\text{Beta}(\alpha_1, \alpha_2)$.

A partire da questa distribuzione possiamo introdurre in modo naturale una generalizzazione k -dimensionale, avente $k + 1$ parametri nel modo seguente.

Definizione 1.2. Consideriamo la funzione di densità di probabilità

$$f(x_1, \dots, x_k) = \frac{\Gamma(\alpha_1 + \dots + \alpha_{k+1})}{\Gamma(\alpha_1) \dots \Gamma(\alpha_{k+1})} x_1^{\alpha_1-1} \dots x_k^{\alpha_k-1} (1 - x_1 - \dots - x_k)^{\alpha_{k+1}-1}$$

che possiamo scrivere in maniera compatta come:

$$f(x_1, \dots, x_k) = \frac{\Gamma\left(\sum_{i=1}^{k+1} \alpha_i\right)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i-1} \cdot \left(1 - \sum_{i=1}^k x_i\right)^{\alpha_{k+1}-1} \mathbf{1}_{\Delta_k} \quad (1.2)$$

in cui $\mathbf{1}_{\Delta_k}$ è la funzione indicatrice del k -simpleso

$$\Delta_k = \left\{ (x_1, \dots, x_k) : x_i \geq 0, i \in \{1, \dots, k\}, \sum_{i=1}^k x_i \leq 1 \right\}$$

e nulla al di fuori, in cui i parametri α_i sono reali positivi per ogni i . Diamo il nome di *distribuzione di Dirichlet (k -dimensionale)* alla distribuzione continua di probabilità avente la (1.2) come densità; denoteremo tale distribuzione come $\text{Dir}(\alpha_1, \dots, \alpha_{k+1})$.

Osservazione 1.1. La generalizzazione è veramente naturale: se prendiamo $k = 1$ nella 1.2 abbiamo esattamente la 1.1, ossia la funzione di densità di probabilità di una beta.

Definizione 1.3. La distribuzione gamma di parametri $a > 0$ e $b > 0$, denotata con $\text{Gamma}(a, b)$, è la distribuzione continua definita dalla funzione di densità

$$f(x) = x^{a-1} \frac{e^{-x/b}}{b^a \Gamma(a)}$$

per $x > 0$. I parametri a e b si dicono rispettivamente parametro di forma e di scala.

La distribuzione gamma è una distribuzione abbastanza comune in statistica, in particolare nella modellizzazione dei tempi d'attesa. Vale il seguente importante risultato.

Proposizione 1.1. *Siano X_1, \dots, X_{k+1} variabili aleatorie indipendenti aventi distribuzioni $\text{Gamma}(\alpha_1, 1), \dots, \text{Gamma}(\alpha_{k+1}, 1)$. Siano inoltre*

$$Y_i = \frac{X_i}{X_1 + \dots + X_{k+1}}, \quad i \in \{1, \dots, k\} \quad (1.3)$$

Allora (Y_1, \dots, Y_k) ha distribuzione di Dirichlet k -dimensionale

$$\text{Dir}(\alpha_1, \dots, \alpha_{k+1}).$$

Possiamo usare questa come definizione alternativa della distribuzione di Dirichlet.

Osservazione 1.2. Per comodità successiva fissiamo sin d'ora la seguente convenzione: generalizziamo i concetti di distribuzione beta e di Dirichlet in modo da consentire che alcuni dei parametri si possano annullare. La variabile aleatoria corrispondente al parametro nullo è degenerare in 0.

Esponiamo ora le principali proprietà delle distribuzioni di Dirichlet.

Proposizione 1.2. Sia (Y_1, \dots, Y_k) un vettore aleatorio avente distribuzione di Dirichlet $\text{Dir}(\alpha_1, \dots, \alpha_{k+1})$; allora per $j < k$ la distribuzione marginale di (Y_1, \dots, Y_j) è la distribuzione di Dirichlet j -dimensionale

$$\text{Dir}(\alpha_1, \dots, \alpha_j, \alpha_{j+1} + \dots + \alpha_{k+1}).$$

Una conseguenza immediata della Proposizione 1.2 è che ciascuna delle Y_i ha come distribuzione marginale una beta di parametri $\alpha_i, \sum \alpha_j - \alpha_i$.

Proposizione 1.3. Sia (Y_1, \dots, Y_k) un vettore aleatorio avente distribuzione di Dirichlet $\text{Dir}(\alpha_1, \dots, \alpha_{k+1})$; allora il vettore aleatorio (Z_1, \dots, Z_j) in cui

$$Z_1 = \sum_{l=1}^{i_1} Y_l, Z_2 = \sum_{l=i_1+1}^{i_2} Y_l, \dots, Z_j = \sum_{l=i_{j-1}+1}^{i_j} Y_l,$$

con gli i_h interi tali che $0 < i_1 < \dots < i_j \leq k$, ha distribuzione di Dirichlet j -dimensionale

$$\text{Dir}\left(\sum_{l=1}^{i_1} \alpha_l, \dots, \sum_{l=i_{j-1}+1}^{i_j} \alpha_l, \sum_{l=i_j+1}^{k+1} \alpha_l\right).$$

DIMOSTRAZIONE. Possiamo vedere le Y_i come nella (1.1), allora

$$Z_1 = \sum_{l=1}^{i_1} Y_l = \frac{X_1 + \dots + X_{i_1}}{X_1 + \dots + X_{k+1}}, \dots, Z_j = \frac{X_{i_{j-1}+1} + \dots + X_{i_j}}{X_1 + \dots + X_{k+1}},$$

con le X_i variabili aleatorie indipendenti di distribuzione Gamma $(\alpha_1, 1)$. Allora, poiché $X_1 + \dots + X_{i_1}$ ha distribuzione Gamma $(\alpha_1 + \dots + \alpha_{i_1}, 1)$, Z_1 ha, per la proposizione 1.1, distribuzione

$$\text{Beta}(\alpha_1 + \dots + \alpha_{i_1}, \alpha_{i_1+1} + \dots + \alpha_{k+1}).$$

Ciò vale, in modo del tutto analogo, per le altre Z_i . \square

Infine abbiamo che, dato un vettore aleatorio (X_1, \dots, X_k) avente distribuzione di Dirichlet $\text{Dir}(\alpha_1, \dots, \alpha_{k+1})$, la variabile aleatoria X_i ha media, varianza e covarianza uguali a

$$\mathbb{E}[X_i] = \frac{\alpha_i}{\alpha_0}, \quad \text{Var}[X_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}, \quad \text{Cov}[X_i X_j] = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}, \quad \text{per } j \neq i.$$

1.2 Teorema di Kolmogorov

Richiamiamo ora, per prima cosa, il concetto di processo stocastico. Supponiamo che \mathcal{X} sia uno spazio polacco.

Definizione 1.4. Dato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, un *processo stocastico* a valori in uno spazio \mathcal{X} , detto *spazio degli stati*, è una famiglia di variabili aleatorie $(X_t)_{t \in T}$ parametrizzate da un insieme T , in cui ciascuna delle X_t è a valori in \mathcal{X} .

Osservazione 1.3. Osserviamo che, per ogni $t \in T$ fissato, abbiamo una variabile aleatoria

$$\omega \rightarrow X_t(\omega), \quad \omega \in \Omega,$$

mentre, se consideriamo $\omega \in \Omega$ fissato, abbiamo la funzione

$$t \rightarrow X_t(\omega), \quad t \in T$$

che prende il nome di *traiettoria* di X_t . A volte, invece, scriviamo $X(t, \omega)$, per sottolineare che si tratta di una mappa da $T \times \Omega$ a \mathcal{X} .

Osservazione 1.4. È importante sottolineare che possiamo identificare ciascun ω con la funzione $t \rightarrow X_t(\omega)$ da T in \mathcal{X} . In quest'ottica possiamo interpretare Ω come un sottoinsieme dello spazio $\widehat{\Omega} = \mathcal{X}^T$ delle funzioni da T in \mathcal{X} . Allora la tribù \mathcal{A} conterrà la tribù \mathcal{C} generata dagli insiemi

$$\{\omega : \omega(t_1) \in A_1, \dots, \omega(t_k) \in A_k\}, \quad A_i \subset \mathcal{X} \text{ boreliani.}$$

Possiamo allora guardare un processo stocastico da questo nuovo punto di vista, interpretandolo come una misura di probabilità sullo spazio (misurabile) $(\mathcal{X}^T, \mathcal{C})$.

Definizione 1.5. Dato un processo stocastico $X = (X_t)_{t \in T}$ a valori in \mathcal{X} su uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, si dicono *distribuzioni finito-dimensionali* di X le misure $\mathbb{P}_{t_1, \dots, t_k}$ ciascuna definita su \mathcal{X}^k al variare di $k \in \mathbb{N}$ da

$$\mathbb{P}_{t_1, \dots, t_k} = \mathbb{P}(X_{t_1}, \dots, X_{t_k})^{-1} \quad t_1, \dots, t_k \in T.$$

Esse costituiscono un sistema proiettivo di misure di probabilità, relativo allo spazio \mathcal{X} .

Proviamo ad invertire questo procedimento: assegnato un sistema proiettivo di misure di probabilità $\{\mathbb{P}_{t_1, \dots, t_k}, k \in \mathbb{N}, t_1, \dots, t_k \in T\}$ relativo allo spazio \mathcal{X} , vorremmo costruire un processo stocastico che abbia le $\mathbb{P}_{t_1, \dots, t_k}$ per distribuzioni finito-dimensionali. Il prossimo teorema, dovuto a Kolmogorov, mostra che ciò è possibile, a patto che la famiglia soddisfi la condizione seguente.

Definizione 1.6. Una famiglia $\{\mathbb{P}_{t_1, \dots, t_k}, k \in \mathbb{N}, t_1, \dots, t_k \in T\}$ di misure di probabilità su \mathcal{X}^k si dice *consistente* (secondo Kolmogorov) se per ogni

$$\{j_1, \dots, j_h\} \subseteq \{1, \dots, k\}$$

si ha la relazione

$$\mathbb{P}_{t_1, \dots, t_k} \big|_{\{t_{j_1}, \dots, t_{j_h}\}} = \mathbb{P}_{t_{j_1}, \dots, t_{j_h}}. \quad (1.4)$$

Teorema 1.4 (Kolmogorov). Sia $\{\mathbb{P}_{t_1, \dots, t_k}, k \in \mathbb{N}, t_1, \dots, t_k \in T\}$ una famiglia consistente di misure di probabilità su \mathcal{X}^k . Allora esistono uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ ed un processo stocastico $(X_t)_{t \in T}$ su Ω a valori in \mathcal{X} tali che, per ogni $t_1, \dots, t_k \in T$ ed ogni $k \in \mathbb{N}$, le $\mathbb{P}_{t_1, \dots, t_k}$ siano le distribuzioni finito dimensionali di $\{X_t\}_{t \in T}$, ossia

$$\mathbb{P}_{t_1, \dots, t_k} = \mathbb{P}(X_{t_1}, \dots, X_{t_k})^{-1}.$$

1.3 Misure aleatorie

Sia \mathcal{X} uno spazio polacco. In \mathcal{X} prendiamo la tribù \mathcal{B} dei boreliani e, al suo interno, l'anello \mathcal{B} costituito da tutti gli insiemi limitati di \mathcal{B} . Indichiamo con $F = F(\mathcal{X})$ la classe delle funzioni \mathcal{B} -misurabili $f : \mathcal{X} \rightarrow \mathbb{R}_+$ e con $\mathcal{C}_0 = \mathcal{C}_0(\mathcal{X})$ la sottoclasse di quelle funzioni di F che son continue e a supporto compatto. Diciamo infine che una classe $C \subset \mathcal{B}$ è un ricoprimento se ogni insieme $B \in \mathcal{B}$ ammette un ricoprimento finito con insiemi di C .

Diciamo che una misura μ su $(\mathcal{X}, \mathcal{B})$ è localmente finita o di Radon se $\mu(B) < \infty$ per ogni $B \in \mathcal{B}$ e denotiamo con $\mathfrak{M} = \mathfrak{M}(\mathcal{X})$ la famiglia di queste misure. Sia poi $\mathcal{M} = \mathcal{M}(\mathcal{X})$ la tribù in \mathfrak{M} generata dalle mappe $\mu \mapsto \mu(B)$, con $B \in \mathcal{B}$.

Date una generica misura $\mu \in \mathfrak{M}$ e una funzione $f \in F$, possiamo definire l'integrale $\mu[f]$ e la misura $f\mu$ nel modo seguente:

$$\mu[f] = \int_{\mathcal{X}} f(s) \mu(ds), \quad (f\mu)(B) = \int_B f(s) \mu(ds) \quad B \in \mathcal{B}.$$

Nel caso particolare in cui la funzione f sia la funzione indicatrice di B , $\mathbf{1}_B$ per qualche $B \in \mathcal{B}$, la misura $f\mu = \mathbf{1}_B\mu$ si dice restrizione di μ a B .

Lemma 1.5. La mappa $\mu \mapsto \mu[f]$ di $(\mathfrak{M}, \mathcal{M})$ in $[0, +\infty]$ è misurabile per ogni $f \in F$. Se $f \in F$ è limitata, allora $\mu \mapsto f\mu$ è una mappa misurabile di $(\mathfrak{M}, \mathcal{M})$ in $(\mathfrak{M}, \mathcal{M})$ e, in particolare, $\mu \mapsto \mathbf{1}_B\mu$ è una mappa misurabile di $(\mathfrak{M}, \mathcal{M})$ in se stesso per ogni $B \in \mathcal{B}$.

DIMOSTRAZIONE. Il primo punto della tesi è banalmente vero per le funzioni semplici per come abbiamo definito \mathcal{M} e segue nel caso più generale per convergenza monotona. Il secondo punto segue dal primo osservando ancora una volta la definizione di \mathcal{M} . \square

Definizione 1.7. Chiamiamo *misura aleatoria* su \mathcal{X} una qualunque funzione misurabile da uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ in $(\mathfrak{M}, \mathcal{M})$.

Osservazione 1.5. Possiamo permettere che una misura aleatoria assuma valori al di fuori di \mathfrak{M} su un insieme $A \in \mathcal{A}$ con $\mathbb{P}(A) = 0$.

Lemma 1.6. *La famiglia delle misure aleatorie su \mathcal{X} è chiusa rispetto all'addizione e alla moltiplicazione per variabili aleatorie a valori in \mathbb{R}_+ . Inoltre una serie $\sum_i \mu_i$ di misure aleatorie è a sua volta una misura aleatoria se e solo se $\sum_i \mu_i(B) < \infty$ quasi certamente per ogni $B \in \mathcal{B}$.*

DIMOSTRAZIONE. Il primo punto della tesi segue immediatamente dalla definizione data di \mathcal{M} e dal fatto che le variabili aleatorie formano una famiglia chiusa rispetto all'addizione ed alla moltiplicazione. Per quanto riguarda il secondo punto, abbiamo per convergenza monotona che $\sum_i \mu_i$ è σ -additiva su \mathcal{B} e quindi dà luogo ad una misura. La necessità della condizione segue dal fatto che $\mu(B) < \infty$ quasi certamente per ogni misura aleatoria μ e ogni $B \in \mathcal{B}$. Viceversa supponiamo che $\sum_i \mu_i(B) < \infty$ quasi certamente, $B \in \mathcal{B}$. Se consideriamo questa disuguaglianza per tutti gli insiemi B appartenenti ad una qualche famiglia ricoprente numerabile, abbiamo che l'insieme \mathbb{P} -trascurabile delle eccezioni può essere preso indipendentemente da B . Allora $\sum_i \mu_i \in \mathfrak{M}$ quasi certamente. La misurabilità di $\sum_i \mu_i$ segue dal fatto che $\sum_i \mu_i(B)$ è una variabile aleatoria per ogni $B \in \mathcal{B}$. \square

Definizione 1.8. La *distribuzione* di una misura aleatoria μ è la misura di probabilità $\mathbb{P}\mu^{-1}$ su $(\mathfrak{M}, \mathcal{M})$ data da

$$(\mathbb{P}\mu^{-1})(M) = \mathbb{P}(\mu^{-1}(M)) = \mathbb{P}(\mu \in M), \quad m \in \mathcal{M}.$$

L'*intensità* di μ è la funzione d'insieme $\mathbb{E}\mu$ data da

$$(\mathbb{E}\mu)(B) = \mathbb{E}[\mu(B)], \quad B \in \mathcal{B},$$

dove $\mathbb{E} = \mathbb{E}^{\mathbb{P}}$ denota la speranza rispetto alla misura di probabilità \mathbb{P} . Osserviamo che $\mathbb{E}\mu$ è sempre una misura, anche se non appartiene necessariamente a \mathfrak{M} .

Sia μ una misura aleatoria su \mathcal{X} ; poniamo

$$\mathbb{P}_{B_1, \dots, B_k} = \mathbb{P}(\mu(B_1), \dots, \mu(B_k))^{-1},$$

al variare di $k \in \mathbb{N}$, $B_1, \dots, B_k \in \mathcal{B}$. Allora la proiezione di $\mathbb{P}_{B_1, \dots, B_k}$ sul sottospazio generato dalle coordinate $\{j_1, \dots, j_h\} \subset \{1, \dots, k\}$ coincide con $\mathbb{P}_{B_{j_1}, \dots, B_{j_h}}$. Come abbiamo già detto nella sezione precedente, una famiglia $\{\mathbb{P}_{B_1, \dots, B_k}\}$ di misure di probabilità che soddisfa tutte le relazioni di questo tipo si dice consistente. Possiamo allora invocare il teorema di Kolmogorov 1.4, che ci garantisce l'esistenza di un processo stocastico ξ su \mathcal{B} tale che

$$\mathbb{P}(\xi(B_1), \dots, \xi(B_k))^{-1} = \mathbb{P}_{B_1, \dots, B_k}, \quad k \in \mathbb{N}, B_1, \dots, B_k \in \mathcal{B}.$$

Per garantire che ξ ammetta una versione μ che sia una misura aleatoria, abbiamo bisogno di alcune ipotesi aggiuntive.

Teorema 1.7. *Sia $\{\mathbb{P}_{B_1, \dots, B_k}\}$, con $k \in \mathbb{N}$ e $B_1, \dots, B_k \in \mathcal{B}$, una famiglia consistente di misure di probabilità. Allora esiste una misura aleatoria μ sullo spazio \mathcal{X} che soddisfa*

$$\mathbb{P}(\mu(B_1), \dots, \mu(B_k))^{-1} = \mathbb{P}_{B_1, \dots, B_k}, \quad k \in \mathbb{N}, B_1, \dots, B_k \in \mathcal{B}$$

se e solo se

- (i) $\mathbb{P}_{A, B, A \cup B}((x, y, z) : x + y = z) = 1$ per tutti gli $A, B \in \mathcal{B}$ disgiunti tra loro,
- (ii) $\mathbb{P}_{B_n} \xrightarrow{w} \delta_0$ per ogni $(B_n)_n$ con $B_n \downarrow \emptyset$, in cui \xrightarrow{w} denota la convergenza debole.

Capitolo 2

Il processo di Wiener

Una possibile costruzione del processo di Wiener, o moto browniano, si basa sul teorema di Kolmogorov enunciato in precedenza. Abbiamo voluto inserire questo capitolo nella tesi non solo perché possiamo presentare in un caso ben noto l'applicazione del teorema di Kolmogorov alla costruzione di un processo stocastico, ma anche perché il processo di Wiener ritornerà più avanti nel capitolo 5 come metro di paragone e come riferimento euristico per la costruzione del processo di Dirichlet.

2.1 Costruzione

Definizione 2.1. Una famiglia \mathcal{I} di variabili aleatorie su uno spazio probabillizzato $(\Omega, \mathcal{A}, \mathbb{P})$ si dice *famiglia gaussiana* se per ogni scelta di $X_1, \dots, X_k \in \mathcal{I}$ e $\gamma_1, \dots, \gamma_k \in \mathbb{R}$ la variabile aleatoria $\gamma_1 X_1 + \dots + \gamma_k X_k$ è normale.

È immediata la caratterizzazione seguente.

Lemma 2.1. \mathcal{I} è una famiglia gaussiana se e solo se per ogni $X_1, \dots, X_k \in \mathcal{I}$ la variabile aleatoria (X_1, \dots, X_k) è normale.

Definizione 2.2. Un processo stocastico $X = (X_t)_{t \in T}$ è un *processo gaussiano* se la famiglia $(X_t)_{t \in T}$ è una famiglia gaussiana.

Per comodità consideriamo d'ora in poi processi stocastici a valori reali con $T = [0, +\infty[= \mathbb{R}_+$ come insieme dei parametri. Osserviamo che chiedere che il processo $X = (X_t)_{t \geq 0}$ sia gaussiano equivale a domandare che, per ogni k -upla $(t_1, \dots, t_k) \subset T$, la variabile aleatoria $Y = (X_{t_1}, \dots, X_{t_k}) \in \mathbb{R}^k$ abbia distribuzione normale, ossia che esistano un vettore $M \in \mathbb{R}^k$ ed una matrice

definita positiva $C = (c_{j,l}) \in \mathbb{R}^{k \times k}$ tali che

$$\mathbb{E} \left[\exp \left(i \sum_{j=1}^k \gamma_j Y_j \right) \right] = \exp \left(-\frac{1}{2} \sum_{j,l} \gamma_j c_{j,l} \gamma_l + i \sum_j \gamma_j M_j \right), \quad (2.1)$$

per ogni $\gamma = (\gamma_1, \dots, \gamma_k) \in \mathbb{R}^k$. Inoltre, se vale la (2.1), $M = \mathbb{E}[Y]$ è la speranza della variabile aleatoria Y e

$$C = (c_{j,l}) = \mathbb{E}[(Y_j - M_j)(Y_l - M_l)]$$

è la matrice di covarianza di Y .

Diamo ora una prima definizione di processo di Wiener (grossolano).

Definizione 2.3. Un processo $(W_t)_{t \geq 0}$ a valori reali relativo ad una filtrazione $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ è un *processo di Wiener (grossolano)* centrato in 0 se

- (1) $W_0 = 0$ quasi certamente;
- (2) per ogni $0 \leq s \leq t$ la variabile aleatoria $W_t - W_s$ è indipendente da \mathcal{F}_s ;
- (3) per ogni $0 \leq s \leq t$ la variabile aleatoria $W_t - W_s$ ha legge $\mathcal{N}(0, t - s)$.

La (2) si può enunciare anche dicendo che W è ad incrementi indipendenti. In particolare dalla (2) segue che la variabile aleatoria $W_t - W_s$ è indipendente da W_u per ogni $u \leq s$.

Proposizione 2.2. *Un processo di Wiener grossolano è un processo gaussiano.*

DIMOSTRAZIONE. Procediamo per induzione: siano $\gamma_1, \dots, \gamma_k \in \mathbb{R}$, $0 \leq t_1 < \dots < t_k$, dobbiamo provare che $\gamma_1 W_{t_1} + \dots + \gamma_k W_{t_k}$ è una variabile aleatoria reale normale.

Il caso $k = 1$ è immediato. Supponiamo ora che l'affermazione sia vera per $k - 1$, allora abbiamo:

$$\begin{aligned} & \gamma_1 W_{t_1} + \dots + \gamma_k W_{t_k} = \\ & = (\gamma_1 W_{t_1} + \dots + (\gamma_{k-1} + \gamma_k) W_{t_{k-1}}) + \gamma_k (W_{t_k} - W_{t_{k-1}}) \end{aligned}$$

che è una variabile aleatoria normale, vista come somma di due variabili aleatorie reali normali indipendenti. \square

Osservazione 2.1. Sempre per induzione si può provare che, in generale, per ogni $0 \leq t_0 < t_1 < \dots < t_k$ le variabili aleatorie reali $W_{t_j} - W_{t_{j-1}}$, con $k = 1, \dots, k$, sono indipendenti.

Osservazione 2.2. Il punto (2) della definizione 2.3 implica anche la seguente proprietà; per ogni $s \geq 0$ la tribù $\mathcal{T}(W_t - W_s, t \geq s)$ è indipendente da \mathcal{F}_s .

Osservazione 2.3. A volte consideriamo un processo di Wiener rispetto a particolari filtrazioni. In generale \mathcal{F} dovrà contenere la tribù $\mathcal{T}(W_s, s \leq t)$, affinché W sia un processo adattato. Se W è un processo di Wiener relativo alla filtrazione \mathcal{F} , allora è un processo di Wiener anche rispetto ad ogni altra filtrazione $\mathcal{G} \subseteq \mathcal{F}$ contenente la filtrazione naturale. In particolare, se \mathcal{F} è la filtrazione naturale, parleremo di processo di Wiener naturale.

Proposizione 2.3. *W è un processo di Wiener grossolano centrato nell'origine, se e solo se valgono le tre condizioni seguenti:*

1. $W_0 = 0$ quasi certamente;
2. per ogni $0 \leq t_1 < \dots < t_k$, $(W_{t_1}, \dots, W_{t_k})$ è una variabile aleatoria normale centrata;
3. $\mathbb{E}[W_s W_t] = s \wedge t$.

DIMOSTRAZIONE. Sia W un processo di Wiener grossolano; la 1. è ovvia, mentre la 2. segue immediatamente dal fatto che un processo di Wiener è un processo gaussiano. La 3. infine segue dal fatto che, se $s \leq t$,

$$\mathbb{E}[W_t W_s] = \mathbb{E}[(W_t - W_s) W_s] + \mathbb{E}[W_s^2] = s = s \wedge t.$$

Viceversa, se valgono le 1., 2. e 3., (1) è ovvio. Se $0 \leq s < t$, $W_t - W_s$ è normale e, dal momento che W_t e W_s sono centrate, tale è anche $W_t - W_s$. Inoltre, poiché

$$\mathbb{E}[(W_t - W_s)^2] = \mathbb{E}[W_t^2] + \mathbb{E}[W_s^2] - 2\mathbb{E}[W_t W_s] = t + s - 2s = t - s,$$

otteniamo che $W_t - W_s$ ha legge $\mathcal{N}(0, t - s)$, cioè vale la (3).

Infine se $h \leq s \leq t$

$$\mathbb{E}[(W_t - W_s) W_h] = t \wedge h - s \wedge h = 0$$

e $W_t - W_s$ è dunque indipendente dalla tribù $\mathcal{T}(W_h, h \leq s)$, da cui (2). \square

Osservazione 2.4. Nella dimostrazione precedente la prima implicazione è vera per ogni processo di Wiener, mentre la seconda ci dà un processo di Wiener naturale.

Dobbiamo però ancora dimostrare che esiste un processo che soddisfa le condizioni date nella definizione 2.3. Per fare ciò, come anticipato, sfrutteremo il teorema di Kolmogorov 1.4: ci occorre allora una famiglia di misure di probabilità $\{\mathbb{P}_{t_1, \dots, t_k}\}$ che soddisfi la (1.4). Tale famiglia deve essere anche opportunamente collegata alle proprietà che un eventuale processo di Wiener avrebbe.

Fissiamo allora $x \in \mathbb{R}$ e definiamo

$$p(t, x, y) = \frac{1}{\sqrt{(2\pi t)^2}} \cdot \exp\left(-\frac{|x - y|^2}{2t}\right),$$

per $y \in \mathbb{R}$ e $t \in]0, +\infty[$. Per $0 \leq t_1 \leq \dots \leq t_k$ possiamo allora definire la misura $\mathbb{P}_{t_1, \dots, t_k}$ su \mathbb{R}^k nel modo seguente:

$$\mathbb{P}_{t_1, \dots, t_k} = \int p(t_1, 0, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \cdots dx_k, \quad (2.2)$$

in cui $dx = dx_1 \cdots dx_k$ denota la misura di Lebesgue. Poniamo, convenzionalmente, $p(0, x, y) = \delta_x(y)$, dove δ_x è la delta di Dirac.

Questa famiglia di misure è consistente: ciò segue dal fatto che, per ogni $t \geq 0$ ed ogni $x \in \mathbb{R}$,

$$\int_{\mathbb{R}} p(t, x, y) dy = 1.$$

Per il teorema di Kolmogorov 1.4 esistono allora uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ ed un processo stocastico $(W_t)_{t \geq 0}$ su Ω tali che le distribuzioni finito-dimensionali di W_t siano date dalla (2.2), cioè:

$$\mathbb{P}(W_{t_1}, \dots, W_{t_k})^{-1} = \int p(t_1, 0, x_1) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \cdots dx_k. \quad (2.3)$$

Dobbiamo ora mostrare che il processo $(W_t)_{t \geq 0}$ così costruito è un processo di Wiener (grossolano) secondo la definizione 2.3. Per fare ciò verifichiamo che soddisfa le condizioni della proposizione 2.3. Osserviamo subito che, grazie alla convenzione $p(0, x, y) = \delta_x(y)$, abbiamo immediatamente $\mathbb{P}(W_0 = 0) = 1$. Proviamo ora che $(W_t)_{t \geq 0}$ è un processo gaussiano: consideriamo $Y = (W_{t_1}, \dots, W_{t_k})$ e mostriamo che soddisfa la (2.1). Calcoliamone il primo membro, usando la (2.3): otteniamo l'uguaglianza

$$\mathbb{E} \left[\exp \left(i \sum_{j=1}^k \gamma_j W_{t_j} \right) \right] = \exp \left(-\frac{1}{2} \sum_{j,l} \gamma_j c_{j,l} \gamma_l + i \sum_j \gamma_j M_j \right),$$

al variare di $\gamma \in \mathbb{R}^k$, con

$$M = \mathbb{E}[Y] = (0, \dots, 0)$$

e con

$$C = \begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_k \end{pmatrix}.$$

Questo ci dice che $(W_t)_{t \geq 0}$ è un processo gaussiano, ma anche che $\mathbb{E}[W_t] = 0$ per ogni $t \geq 0$ e che

$$\mathbb{E}[W_t^2] = t, \quad \mathbb{E}[W_t W_s] = t \wedge s.$$

Allora abbiamo che il processo $(W_t)_{t \geq 0}$ ottenuto è proprio un processo di Wiener grossolano.

È importante notare che il processo di Wiener ottenuto dal teorema di Kolmogorov non è unico: esistono diverse quadruple $(W_t, \Omega, \mathcal{A}, \mathbb{P})$ che soddisfano la (2.3). Introduciamo allora il concetto di versione.

Definizione 2.4. Consideriamo due processi stocastici $(X_t)_t$ e $(Y_t)_t$ sul medesimo spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$. Si dice che $(X_t)_t$ è una *versione* di $(Y_t)_t$ se

$$\mathbb{P}\{X_t(\omega) = Y_t(\omega)\} = 1$$

per ogni t .

Osservazione 2.5. Se $(X_t)_t$ è una versione di $(Y_t)_t$, allora i due processi X, Y hanno le stesse distribuzioni finito-dimensionali. Pertanto, se vediamo un processo stocastico come una legge di probabilità su $\mathbb{R}^{[0, +\infty[}$, possiamo identificarli anche se le loro traiettorie hanno caratteristiche differenti.

Possiamo allora cercare una versione del processo di Wiener che abbia traiettorie continue. La sua esistenza è garantita dal seguente criterio, dovuto sempre a Kolmogorov.

Teorema 2.4 (Criterio di hölderianità di Kolmogorov). *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, \mathbb{P})$ sia dato un processo reale $(X_t)_t$ a valori in \mathbb{R} . Siano inoltre α, δ, C costanti reali strettamente positive, tali che si abbia*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq C |t - s|^{1+\delta} \tag{2.4}$$

per ogni s, t in \mathbb{R}_+ . Allora esiste una versione del processo a traiettorie localmente hölderiane di esponente γ , per ogni numero reale γ strettamente compreso tra 0 e δ/α .

Nel caso del processo di Wiener $(W_t)_{t \geq 0}$ abbiamo, per $t > s$ ed ogni n ,

$$\begin{aligned} \mathbb{E} [|W_t - W_s|^{2n}] &= \frac{1}{\sqrt{2\pi(t-s)}} \int_{\mathbb{R}} x^{2n} \exp\left(-\frac{x^2}{2(t-s)}\right) dx \\ &= (t-s)^n \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} y^{2n} \exp\left(-\frac{y^2}{2}\right) dy \\ &= C_n |t-s|^n. \end{aligned}$$

Allora siamo nelle ipotesi del teorema 2.4, con la (2.4) soddisfatta per $\alpha = 2n$, $\delta = n - 1$: esiste dunque una versione del processo di Wiener a traiettorie γ -hölderiane, per $0 < \gamma < \frac{n-1}{2n}$. Inoltre, per la genericità di n , possiamo affermare che ciò vale per ogni γ positivo minore di $1/2$. Possiamo allora affermare che il processo di Wiener ammette una versione a traiettorie continue.

2.2 Misura di Wiener

Consideriamo lo spazio \mathcal{C}_0 delle funzioni continue da $[0, +\infty[$ in \mathbb{R} che valgano 0 nell'origine e il processo di Wiener a traiettorie continue $(W_t)_{t \in [0, +\infty[}$ a valori in \mathbb{R} . Allora la legge del processo di Wiener induce una misura \mathbb{P}^W sullo spazio \mathcal{C}_0 dotato della tribù \mathcal{F} ottenuta prendendo, sullo spazio prodotto $\mathbb{R}^{\mathbb{R}}$ di tutte le applicazioni di \mathbb{R} in \mathbb{R} , il prodotto di una famiglia di copie della tribù boreliana di \mathbb{R} , prendendone poi la traccia sull'insieme \mathcal{C}_0 (non appartenente alla tribù prodotto). Tale misura è detta *misura di Wiener*. Un'analogia costruzione si può fare nel caso $T = [0, 1]$, considerando lo spazio \mathcal{G}^* di tutte le mappe continue $g : [0, 1] \rightarrow \mathbb{R}$ con $g(0) = 0$. Anche in questo caso la misura \mathbb{P}^W indotta dal processo di Wiener prende il nome di *misura di Wiener*.

Possiamo interpretare e costruire la misura di Wiener \mathbb{P}^W sullo spazio \mathcal{G}^* sfruttando il seguente approccio euristico, dovuto a Feynman:

$$\mathbb{P}^W(dg) = \frac{1}{Z} e^{-H(g)} \mathbb{P}^{\mathcal{G}^*}(dg), \quad (2.5)$$

in cui $H(g)$ è il funzionale dell'energia

$$H(g) = \frac{1}{2} \int_0^1 g'(t)^2 dt,$$

mentre $\mathbb{P}^{\mathcal{G}^*}(dg)$ è una "distribuzione uniforme" sullo spazio \mathcal{G}^* di tutte le mappe continue $g : [0, 1] \rightarrow \mathbb{R}$ con $g(0) = 0$. Una tale distribuzione uniforme non può esistere, perché non esiste un concetto di misura invariante per traslazioni in uno spazio di dimensione infinita. Inoltre, anche se una distribuzione siffatta

esistesse, quasi tutte le mappe g avrebbero energia infinita; possiamo però risolvere questo problema nel modo seguente.

Data una qualunque partizione finita $\{0 = t_0 < \dots < t_N = 1\}$ di $[0, 1]$, possiamo sostituire all'energia $H(g)$ della mappa g l'energia della spezzata che interpola g :

$$H_N(g) = \inf \{H(\tilde{g}) : \tilde{g} \in \mathcal{G}^*, \tilde{g}(t_i) = g(t_i) \forall i\} = \sum_{i=1}^N \frac{|g(t_i) - g(t_{i-1})|^2}{2(t_i - t_{i-1})}.$$

Ora, prendendo ispirazione dalla (2.5), abbiamo la seguente rappresentazione esplicita per le leggi finito dimensionali:

$$\begin{aligned} \mathbb{P}^W(g_{t_1} \in dx_1, \dots, g_{t_N} \in dx_N) &= \\ &= \frac{1}{Z_N} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right) p_N(dx_1, \dots, dx_N), \end{aligned} \quad (2.6)$$

in cui $p_N(dx_1, \dots, dx_N) = \mathbb{P}^{\mathcal{G}^*}(g_{t_1} \in dx_1, \dots, g_{t_N} \in dx_N)$ è una “distribuzione uniforme” su \mathbb{R}^N e Z_N una costante di normalizzazione. Se prendiamo come p_N la misura N -dimensionale di Lebesgue, abbiamo che il secondo membro della (2.6) è una famiglia proiettiva di misure di probabilità. Essa ammette un unico limite proiettivo¹, che è proprio la misura di Wiener \mathbb{P}^W su \mathcal{G}^* .

¹Una misura di probabilità sulla tribù prodotto è *limite proiettivo* di un sistema proiettivo se ne prolunga tutte le misure di probabilità.

Capitolo 3

Il processo di Dirichlet

Arriviamo ora ad introdurre l'oggetto centrale di questa tesi. Si tratta di un processo stocastico nel senso della definizione 1.4, ma con la peculiarità, come vedremo, di avere come insieme dei parametri una tribù. Questo ci permette di interpretare ciascuna traiettoria del processo come misura (di probabilità) sullo spazio su cui il processo è definito e il processo stesso come famiglia aleatoria di misure, ossia come misura aleatoria. Sottolineiamo infine, prima di iniziare con la trattazione, che lavoriamo su uno spazio \mathcal{X} qualunque, fatto non vero per ogni costruzione del processo di Dirichlet, come vedremo nel successivo capitolo 4.

Diamo per prima cosa la definizione. Essa ci farà da guida nella prima parte del capitolo, finché non arriveremo a giustificarla pienamente.

Definizione 3.1. Dati un insieme \mathcal{X} ed una tribù \mathcal{A} su \mathcal{X} , sia α una misura finita, non negativa e non ovunque nulla su $(\mathcal{X}, \mathcal{A})$. Diciamo che μ è un *processo di Dirichlet* di parametro α su \mathcal{X} se è un processo stocastico avente \mathcal{A} come insieme dei parametri, a valori in $[0, +\infty]$ tale che per ogni k ed ogni partizione misurabile (B_1, \dots, B_k) su \mathcal{X} , la legge finito-dimensionale μ_{B_1, \dots, B_k} è la distribuzione di Dirichlet $\text{Dir}(\alpha(B_1), \dots, \alpha(B_k))$, cioè ha densità

$$f(x_1, \dots, x_{k-1}) = \frac{\Gamma\left(\sum_{i=1}^{k+1} \alpha_i\right)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i)} x_1^{\alpha_1-1} \dots x_{k-1}^{\alpha_{k-1}-1} \left(1 - \sum_{i=1}^{k-1} x_i\right)^{\alpha_k-1} \mathbf{1}_\Delta$$

in cui $\alpha_i = \alpha(B_i)$.

Abbiamo dichiarato il nostro obiettivo, ora dobbiamo mostrare che un tale processo esiste. Per fare ciò ci appoggeremo al teorema di Kolmogorov, per via della caratterizzazione voluta delle distribuzioni finito-dimensionali. Siano \mathcal{X} un insieme ed \mathcal{A} una tribù su \mathcal{X} . Notiamo che in questo caso lo spazio dei

parametri è la tribù \mathcal{A} . Questo richiede un po' di attenzioni nell'assegnare le distribuzioni finito-dimensionali. Consideriamo le partizioni finite misurabili di \mathcal{X} , ossia le famiglie finite (B_1, \dots, B_k) che sono una partizione di \mathcal{X} , in cui $B_i \in \mathcal{A}$ per ogni i e $B_i \cap B_j = \emptyset$ per $i \neq j$. Supponiamo di avere, per ogni partizione misurabile finita (B_1, \dots, B_k) al variare di $k \in \mathbb{N}$, una distribuzione finito-dimensionale

$$\nu_{B_1, \dots, B_k}.$$

Scriviamo anche, con qualche abuso di notazione, per comodità successiva:

$$(\nu_{B_1}, \dots, \nu_{B_k}) = \nu_{B_1, \dots, B_k}.$$

Osserviamo che, se abbiamo le distribuzioni finito-dimensionali associate alle partizioni misurabili, possiamo ricavarle anche per arbitrarie famiglie finite (A_1, \dots, A_j) di insiemi \mathcal{A} -misurabili. Data una famiglia finita qualunque di insiemi misurabili (A_1, \dots, A_j) , costruiamo i $k = 2^j$ insiemi che si ottengono dalle intersezioni tra gli A_i e i loro complementari, cioè gli insiemi B_{m_1, \dots, m_j} con gli m_i a valori in $\{0, 1\}$ nel modo seguente:

$$B_{m_1, \dots, m_j} = \bigcap_{i=1}^j A_i^{m_i} \quad (3.1)$$

con $A_i^1 = A_i$ e $A_i^0 = A_i^c$. In questo modo otteniamo una partizione misurabile dello spazio \mathcal{X} mediante i B_{m_1, \dots, m_j} . Ora, per ipotesi, abbiamo già la legge finito-dimensionale associata alla partizione misurabile

$$(B_{m_1, \dots, m_j}, m_i \in \{0, 1\}, i = 1, \dots, j). \quad (3.2)$$

Possiamo allora definire la distribuzione finito-dimensionale per (A_1, \dots, A_j) a partire da quella della (3.2) ponendo, per $i \in \{1, \dots, j\}$

$$\nu_{A_i} = \sum \nu_{B_{m_1, \dots, m_j}}, \quad (3.3)$$

in cui la somma è fatta sulle stringhe (m_1, \dots, m_j) in cui $m_i = 1$. Osserviamo che se (A_1, \dots, A_j) è una partizione misurabile, la definizione della distribuzione ad essa associata che si ottiene dalla (3.3) coincide con quella assegnata, a patto di porre

$$\mu_\emptyset = 0. \quad (3.4)$$

In questo modo abbiamo, a partire dalle distribuzioni associate alle partizioni misurabili, quelle associate a famiglie qualunque di insiemi misurabili.

Ricordiamo che le condizioni di consistenza di Kolmogorov (1.4) devono essere verificate su ogni famiglia di insiemi misurabili. Vogliamo ora mostrare che, in realtà, ci basta in questo contesto verificarle sulle sole partizioni misurabili. Enunciamo allora la seguente condizione di compatibilità.

Definizione 3.2. Supponiamo di avere una famiglia di distribuzioni $\{\nu_{B_1, \dots, B_k}\}$ per ogni partizione misurabile (B_1, \dots, B_k) di \mathcal{X} al variare di $k \in \mathbb{N}$. Tale famiglia si dice *compatibile* se, per ogni coppia di partizioni misurabili (B_1, \dots, B_k) e $(B'_1, \dots, B'_{k'})$ in cui la seconda sia raffinamento della prima con

$$B_1 = \bigcup_{i=1}^{r_1} B'_i, \quad B_2 = \bigcup_{i=r_1+1}^{r_2} B'_i, \quad \dots, \quad B_k = \bigcup_{i=r_{k-1}+1}^{k'} B'_i,$$

la distribuzione

$$\left(\sum_{i=1}^{r_1} \nu_{B'_i}, \quad \sum_{i=r_1+1}^{r_2} \nu_{B'_i}, \quad \dots, \quad \sum_{i=r_{k-1}+1}^{k'} \nu_{B'_i} \right)$$

determinata a partire dalla $\nu_{B'_1, \dots, B'_{k'}}$ coincide con la μ_{B_1, \dots, B_k} .

Come vedremo nel lemma seguente, se una famiglia di distribuzioni è definita come sopra e soddisfa la condizione di compatibilità appena enunciata, allora soddisfa la condizione di consistenza di Kolmogorov (1.4).

Lemma 3.1. *Assegnato un sistema di distribuzioni di $\{\nu_{B_1, \dots, B_k}\}$ per ogni partizione misurabile (B_1, \dots, B_k) al variare di $k \in \mathbb{N}$, tale che soddisfi la condizione di compatibilità 3.2 e se, per famiglie qualunque di insiemi misurabili (A_1, \dots, A_j) le distribuzioni $\{\nu_{A_1, \dots, A_j}\}$ sono definite come nelle (3.1), (3.2) e (3.3), allora la famiglia di distribuzioni $\{\nu_{A_1, \dots, A_j}\}$, al variare delle famiglie qualunque di insiemi misurabili (A_1, \dots, A_j) di $(\mathcal{X}, \mathcal{A})$ è consistente secondo la condizione di Kolmogorov 1.6.*

DIMOSTRAZIONE. Dal momento che $\mathcal{X} \cup \emptyset = \mathcal{X}$, abbiamo dalla condizione di compatibilità che $\nu_\emptyset = 0$ q.c. e, di conseguenza, che la distribuzione ν_{A_1, \dots, A_j} è ben definita dalle (3.2) e (3.3). Per verificare che valgano le condizioni di consistenza di Kolmogorov (1.4) dobbiamo mostrare che per ogni j e per (A_1, \dots, A_j) insiemi misurabili qualunque, la distribuzione $\nu_{A_1, \dots, A_{j-1}}$ ottenuta dalla legge ν_{A_1, \dots, A_j} è uguale alla distribuzione $\nu_{A_1, \dots, A_{j-1}}$ ottenuta direttamente dalla definizione.

La legge $(\nu_{A_1}, \dots, \nu_{A_{j-1}})$ ricavata dalla distribuzione $(\nu_{A_1}, \dots, \nu_{A_j})$ coincide con la

$$\left(\sum \nu_{B_{m_1, \dots, m_j}}, \dots, \sum \nu_{B_{m_1, \dots, m_j}} \right), \quad (3.5)$$

in cui le somme sono sulle stringhe (m_1, \dots, m_j) per cui $m_i = 1$, al variare di $i = 1, \dots, j-1$, ottenuta dalla (3.3). La distribuzione $(\nu_{A_1}, \dots, \nu_{A_{j-1}})$ è definita invece come

$$\left(\sum \nu_{B_{m_1, \dots, m_{j-1}}}, \dots, \sum \nu_{B_{m_1, \dots, m_{j-1}}} \right) \quad (3.6)$$

in cui le somme sono sulle stringhe (m_1, \dots, m_{j-1}) per cui $m_i = 1$, al variare di $i = 1, \dots, j-1$, ottenuta dalla legge

$$\left(\nu_{B_{m_1, \dots, m_{j-1}}}, m_i \in \{0, 1\}, i \in \{1, \dots, j-1\} \right)$$

in cui

$$B_{m_1, \dots, m_{j-1}} = \bigcap_{i=1}^{j-1} A_i^{m_i}.$$

Poiché $B_{m_1, \dots, m_{j-1}} = B_{m_1, \dots, m_{j-1}, 1} \cup B_{m_1, \dots, m_{j-1}, 0}$, dalla condizione di compatibilità abbiamo che la distribuzione

$$\left(\nu_{B_{m_1, \dots, m_{j-1}}}, m_i \in \{0, 1\}, i \in \{1, \dots, j-1\} \right)$$

è uguale alla

$$\left(\nu_{B_{m_1, \dots, m_{j-1}, 1}} + \nu_{B_{m_1, \dots, m_{j-1}, 0}}, m_i \in \{0, 1\}, i \in \{1, \dots, j-1\} \right)$$

ottenuta a partire dalla (3.2). Dunque la (3.6) può essere ricavata dalla (3.2) sostituendo $\nu_{B_{m_1, \dots, m_{j-1}}}$ con $\nu_{B_{m_1, \dots, m_{j-1}, 1}} + \nu_{B_{m_1, \dots, m_{j-1}, 0}}$. Con questo accorgimento la (3.6) è formalmente uguale alla (3.5), da cui la tesi. \square

Grazie a questo risultato possiamo allora costruire un processo stocastico avente \mathcal{A} come insieme dei parametri limitandoci ad assegnare una famiglia di distribuzioni $\{\nu_{B_1, \dots, B_k}\}$ per ogni partizione misurabile (B_1, \dots, B_k) al variare di $k \in \mathbb{N}$ che soddisfi la condizione di compatibilità 3.2. Il lemma 3.1 ci garantisce infatti che la famiglia costruita a partire da essa soddisfa le condizioni di consistenza di Kolmogorov. Siamo dunque nelle ipotesi del teorema 1.4 e possiamo determinare, a partire dalle distribuzioni associate alle partizioni misurabili, un processo stocastico.

La definizione 3.1 è allora ben data. Infatti la condizione di compatibilità per le $\{\mu_{B_1, \dots, B_k}\}$ è proprio la proprietà enunciata nella Proposizione 1.3. Allora, per il lemma 3.1 e il teorema di Kolmogorov 1.4, il processo di Dirichlet è ben definito. Osserviamo che il processo di Dirichlet è un processo stocastico su uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$ dato dal teorema di Kolmogorov, avente \mathcal{A} come spazio dei parametri e a valori in $[0, +\infty]$.

Ora possiamo prendere un punto di vista molto speciale, rispetto a questo processo: consideriamo infatti la traiettoria di μ . Essa è, per $\omega \in \Omega$ fissato, una funzione

$$A \in \mathcal{A} \mapsto \mu_A(\omega)$$

che possiamo anche vedere come funzione di A con la scrittura

$$A \in \mathcal{A} \mapsto \mu(A),$$

lasciando cadere ω . Tale funzione è a valori reali, moralmente è una misura. Diciamo moralmente, perché dobbiamo ancora provarlo, cosa che faremo mediante il criterio di esistenza delle misure aleatorie 1.7. D'ora in poi utilizzeremo anche la notazione $\mu(A)$ per μ_A , per mettere in evidenza il punto di vista della misura.

Le proposizioni che andiamo ad affrontare ora ci mostrano lo stretto legame che intercorre tra le proprietà del processo μ e quelle della misura parametro α .

Proposizione 3.2. *Siano μ un processo di Dirichlet su $(\mathcal{X}, \mathcal{A})$ di parametro α e $A \in \mathcal{A}$. Se $\alpha(A) = 0$ allora $\mu(A) = 0$ q.c.; se $\alpha(A) > 0$ allora $\mu(A) > 0$ q.c. Inoltre $\mathbb{E}[\mu(A)] = \frac{\alpha(A)}{\alpha(\mathcal{X})}$, in cui la speranza è rispetto alla misura di probabilità \mathbb{P} .*

DIMOSTRAZIONE. La dimostrazione è immediata, osservando che data la partizione (A, A^c) si ha che $\mu(A)$ ha distribuzione Beta $(\alpha(A), \alpha(A^c))$, in cui consideriamo la versione generalizzata della beta, per la quale è ammesso che un parametro si annulli. \square

Vale la pena sottolineare che $\mu(\mathcal{X}) = 1$ quasi certamente, come conseguenza della proposizione appena vista.

Proposizione 3.3. *Sia μ un processo di Dirichlet su $(\mathcal{X}, \mathcal{A})$ di parametro α . Allora per una fissata successione decrescente di insiemi misurabili $A_n \downarrow \emptyset$ abbiamo $\mu(A_n) \rightarrow 0$ quasi certamente.*

DIMOSTRAZIONE. Dal momento che $A_n \downarrow \emptyset$, $\alpha(A_n) \rightarrow 0$. Allora esiste una sottosuccessione A_{n_j} tale che $\sum \alpha(A_{n_j}) < \infty$. Fissato $\varepsilon > 0$,

$$\sum_{j=1}^{\infty} \mathbb{P}(\mu(A_{n_j}) > \varepsilon) \leq \sum_{j=1}^{\infty} \varepsilon^{-1} \mathbb{E}[\mu(A_{n_j})] = \varepsilon^{-1} \sum_{j=1}^{\infty} \frac{\alpha(A_{n_j})}{\alpha(\mathcal{X})} < \infty.$$

Ora, per il lemma di Borel-Cantelli,

$$\mathbb{P}(\mu(A_{n_j}) > \varepsilon \text{ infinite volte}) = 0$$

e abbiamo quindi provato che $\mu(A_{n_j}) \rightarrow 0$ quasi certamente. Per completare la dimostrazione basta osservare che $\mu(A_n) > \mu(A_{n+1})$ quasi certamente per ogni n . \square

Teorema 3.4. *Il processo di Dirichlet μ di parametro α definito nella definizione 3.1 è una misura aleatoria, nel senso della definizione 1.7.*

DIMOSTRAZIONE. È sufficiente dimostrare che le due ipotesi del teorema 1.7 sono soddisfatte. Osserviamo che la seconda ipotesi è verificata dal processo

di Dirichlet grazie alla proposizione 3.3. D'altra parte il fatto che la prima ipotesi sia soddisfatta è immediato dalle definizioni delle distribuzioni finito-dimensionali date nelle (3.1), (3.2) e (3.3). Allora μ è una misura aleatoria su $(\mathcal{X}, \mathcal{A})$. \square

Osservazione 3.1. Notiamo inoltre che, per quanto osservato dopo la proposizione 3.2, $\mu(\mathcal{X}) = 1$ quasi certamente, quindi possiamo parlare di *misura aleatoria di probabilità*.

Possiamo allora vedere il processo di Dirichlet come una famiglia di parametro aleatorio ω di misure di probabilità su $(\mathcal{X}, \mathcal{A})$. Ciascuna traiettoria di μ è una misura di probabilità su $(\mathcal{X}, \mathcal{A})$.

Osservazione 3.2. Può sembrare che la proposizione 3.2 affermi che α e ciascuna traiettoria di μ abbiano gli stessi insiemi di misura nulla, ossia che α e ogni traiettoria μ siano assolutamente continue l'una rispetto all'altra. In realtà ciò è falso: vedremo infatti più avanti che ogni traiettoria di μ è quasi certamente una distribuzione discreta. Pertanto α e ogni traiettoria di μ possono essere singolari l'una rispetto all'altra. L'accortezza è osservare che l'insieme trascurabile fuori dal quale vale la proposizione 3.2 può dipendere da A .

Proposizione 3.5. *Siano μ un processo di Dirichlet su $(\mathcal{X}, \mathcal{A})$ di parametro α e ν una misura di probabilità assegnata su $(\mathcal{X}, \mathcal{A})$ assolutamente continua rispetto ad α . Allora, per ogni $k \in \mathbb{N}$, A_1, \dots, A_k insiemi misurabili ed $\varepsilon > 0$,*

$$\mathbb{P}(|\mu(A_i) - \nu(A_i)| < \varepsilon \text{ per } i = 1, \dots, k) > 0.$$

DIMOSTRAZIONE. Costruiamo B_{h_1, \dots, h_k} come nella (3.1) e osserviamo che

$$\begin{aligned} & \mathbb{P}(|\mu(A_i) - \nu(A_i)| < \varepsilon \text{ per } i = 1, \dots, k) \geq \\ & \geq \mathbb{P}\left(\sum |\mu(B_{h_1, \dots, h_k}) - \nu(B_{h_1, \dots, h_k})| < \varepsilon \text{ per } i = 1, \dots, k\right), \end{aligned}$$

in cui la somma è fatta su tutte le stringhe (h_1, \dots, h_k) in cui $h_i = 1$, al variare di $i = 1, \dots, k$. A questo punto è sufficiente mostrare che

$$\mathbb{P}(|\mu(B_{h_1, \dots, h_k}) - \nu(B_{h_1, \dots, h_k})| < 2^{-k}\varepsilon \forall (h_1, \dots, h_k)) > 0.$$

Se $\alpha(B_{h_1, \dots, h_k}) = 0$, allora anche $\nu(B_{h_1, \dots, h_k}) = 0$ e $\mu(B_{h_1, \dots, h_k}) = 0$ quasi certamente. D'altra parte per quelle stringhe (h_1, \dots, h_k) per cui

$$\alpha(B_{h_1, \dots, h_k}) > 0,$$

la legge della corrispondente $\mu(B_{h_1, \dots, h_k})$ dà peso positivo a tutti gli aperti nell'insieme

$$\sum \mu(B_{h_1, \dots, h_k}) = 1$$

in cui la somma è fatta sulle stringhe (h_1, \dots, h_k) tali che $\alpha(B_{h_1, \dots, h_k}) > 0$, fatto che conclude la dimostrazione, per l'assoluta continuità di ν rispetto ad α . \square

Proposizione 3.6. *Sia μ un processo di Dirichlet di parametro α su $(\mathcal{X}, \mathcal{A})$, sia inoltre X un campione estratto da μ . Allora, per $A \in \mathcal{A}$,*

$$\mathbb{P}(X \in A) = \frac{\alpha(A)}{\alpha(\mathcal{X})}.$$

DIMOSTRAZIONE. Dal momento che $\mathbb{P}(X \in A | \mu_A) = \mu_A$ quasi certamente, abbiamo, per la proposizione 3.2,

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{E}[\mathbb{P}(X \in A | \mu(A))] \\ &= \mathbb{E}[\mu(A)] \\ &= \frac{\alpha(A)}{\alpha(\mathcal{X})}, \end{aligned}$$

fatto che completa la dimostrazione. \square

Proposizione 3.7. *Sia μ un processo di Dirichlet di parametro α su $(\mathcal{X}, \mathcal{A})$, sia inoltre X un campione estratto da μ . Sia (B_1, \dots, B_k) una partizione misurabile di \mathcal{X} , e sia $A \in \mathcal{A}$. Allora,*

$$\begin{aligned} &\mathbb{P}\{X \in A, \mu(B_1) \leq y_1, \dots, \mu(B_k) \leq y_k\} = \\ &= \sum_{j=1}^k \frac{\alpha(B_j \cap A)}{\alpha(\mathcal{X})} D(y_1, \dots, y_k | \alpha_1^{(j)}, \dots, \alpha_k^{(j)}), \end{aligned} \quad (3.7)$$

in cui $D(y_1, \dots, y_k | \alpha_1^{(j)}, \dots, \alpha_k^{(j)})$ è la funzione di densità della distribuzione di Dirichlet $\text{Dir}(\alpha_1, \dots, \alpha_k)$ e

$$\alpha_i^{(j)} = \begin{cases} \alpha(B_i) & \text{se } i \neq j \\ \alpha(B_i) + 1 & \text{se } i = j \end{cases}.$$

DIMOSTRAZIONE. Poniamo $B_{j,1} = B_j \cap A$ e $B_{j,0} = B_j \cap A^c$ per $j = 1, \dots, k$. Sia inoltre $Y_{j,l} = \mu(B_{j,l})$ per $j = 1, \dots, k$ e $l = 0, 1$. Abbiamo

$$\mathbb{P}(X \in A | Y_{j,l}, j = 1, \dots, k, l = 0, 1) = \sum_{j=1}^k Y_{j,1} \quad \text{q.c.} \quad (3.8)$$

Allora per $y_{j,l} \in [0, 1]$ qualunque, con $j = 1, \dots, k$ e $l = 0, 1$ possiamo ottenere

$$\mathbb{P}(X \in A, Y_{j,l} \leq y_{j,l}, j = 1, \dots, k, l = 0, 1)$$

integrando la (3.8) rispetto alla legge delle $Y_{j,l}$ sull'insieme

$$\{Y_{j,l} \leq y_{j,l}, j = 1, \dots, k, l = 0, 1\}.$$

Quindi, per le proprietà della distribuzione di Dirichlet, l'integrale è uguale a

$$\sum_{j=1}^k \frac{\alpha(B_{j,1})}{\alpha(\mathcal{X})} D\left(\mathbf{y} | \alpha_{1,0}^{(j)}, \dots, \alpha_{k,0}^{(j)}, \alpha_{1,1}^{(j)}, \dots, \alpha_{k,1}^{(j)}\right)$$

in cui

$$\mathbf{y} = (y_{1,0}, \dots, y_{k,0}, y_{1,1}, \dots, y_{k,1})$$

e in cui

$$\alpha_{i,l}^{(j)} = \begin{cases} \alpha(B_{i,l}) & \text{se } i \neq j \\ \alpha(B_{i,l}) + 1 & \text{se } i = j. \end{cases}$$

La tesi della proposizione segue dunque da questo fatto, sfruttando la proprietà delle distribuzioni di Dirichlet vista nella proposizione 1.3, dal momento che $\mu(B_j) = Y_{j,0} + Y_{j,1}$ quasi certamente. \square

Prima di concludere questo capitolo vediamo un'ultima importante proprietà del processo di Dirichlet. La distribuzione condizionale di un processo di Dirichlet μ , assegnato un campione (X_1, \dots, X_n) estratto da μ , è a sua volta un processo di Dirichlet.

Per $x \in \mathcal{X}$ denotiamo con δ_x la misura su $(\mathcal{X}, \mathcal{A})$ che dà massa 1 al punto x :

$$\delta_x(A) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A. \end{cases}$$

Teorema 3.8. *Sia μ un processo di Dirichlet di parametro α su $(\mathcal{X}, \mathcal{A})$ e sia (X_1, \dots, X_n) un campione di lunghezza n estratto da μ . Allora la distribuzione condizionale di μ dato (X_1, \dots, X_n) è quasi certamente un processo di Dirichlet di parametro*

$$\alpha + \sum_{i=1}^n \delta_{X_i}.$$

DIMOSTRAZIONE. Osserviamo innanzitutto che ci basta dimostrare il teorema nel caso $n = 1$, perché il risultato più generale segue poi per induzione applicando più volte il caso $n = 1$. Sia (B_1, \dots, B_k) una partizione misurabile di \mathcal{X} e sia $A \in \mathcal{A}$. Si verifica facilmente che le leggi marginali di una legge condizionale di un processo sono identiche alle leggi condizionali delle

marginali. Di conseguenza dobbiamo far vedere che la legge condizionale di $\mu(B_1), \dots, \mu(B_k)$ dato X , un campione estratto da μ , ha come legge

$$D(y_1, \dots, y_k | \alpha(B_1) + \delta_X(B_1), \dots, \alpha(B_k) + \delta_X(B_k)). \quad (3.9)$$

Possiamo fare ciò mostrando che l'integrale di (3.9) rispetto alla legge marginale di X su A è uguale alla probabilità (3.7). Usando la legge marginale di X come ottenuto nella proposizione 3.6, abbiamo

$$\begin{aligned} & \int_A D(y_1, \dots, y_k | \alpha(B_1) + \delta_x(B_1), \dots, \alpha(B_k) + \delta_x(B_k)) d\alpha(x) / \alpha(\mathcal{X}) \\ &= \sum_{j=1}^k \int_{B_j \cap A} D(y_1, \dots, y_k | \alpha_1^{(j)}, \dots, \alpha_k^{(j)}) d\alpha(x) / \alpha(\mathcal{X}) \\ &= \sum_{j=1}^k \frac{\alpha(B_j \cap A)}{\alpha(\mathcal{X})} D(y_1, \dots, y_k | \alpha_1^{(j)}, \dots, \alpha_k^{(j)}), \end{aligned}$$

uguaglianza che completa la dimostrazione. \square

Osservazione 3.3. Quest'ultima proposizione ci dà un cenno di una importantissima proprietà del processo di Dirichlet, il fatto che, visto come famiglia di distribuzioni di probabilità, si concentri su distribuzioni discrete. Dimostrare questa proprietà a partire dalla definizione 3.1 non è semplicissimo, preferiamo quindi rimandarla al capitolo successivo, in cui sarà conseguenza immediata delle definizioni alternative equivalenti che daremo.

Capitolo 4

Altri approcci

In questo capitolo vedremo alcuni approcci al processo di Dirichlet alternativi rispetto a quello dovuto a Ferguson, visto nel capitolo 3. Questi nuovi punti di vista ci danno la possibilità di ricavare facilmente alcune delle proprietà caratteristiche del processo di Dirichlet, come il suo essere concentrato sulle distribuzioni discrete di probabilità. Ci permettono inoltre di collegare il processo di Dirichlet ad altri oggetti noti e di darne rappresentazioni intuitivamente più immediate.

4.1 Urne di Pólya

Questo approccio al processo di Dirichlet è dovuto a Blackwell e McQueen, che l'hanno presentato nel loro articolo [BM73].

In questa sezione consideriamo \mathcal{X} uno spazio separabile metrico completo. Sia inoltre α una misura finita positiva sui boreliani di \mathcal{X} . Ricordiamo che una misura aleatoria di probabilità μ su \mathcal{X} è un processo di Dirichlet se per ogni partizione finita (B_1, \dots, B_k) di \mathcal{X} il vettore $(\mu_{B_1}, \dots, \mu_{B_k})$ ha legge Dirichlet di parametri $(\alpha(B_1), \dots, \alpha(B_k))$.

Diciamo che una successione di variabili aleatorie $\{X_n, n \geq 1\}$ su un opportuno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ a valori in \mathcal{X} è una successione di Pólya di parametro α se per ogni $B \subset \mathcal{X}$ valgono

$$\mathbb{P}(X_1 \in B) = \frac{\alpha(B)}{\alpha(\mathcal{X})} \quad (4.1)$$

e

$$\mathbb{P}(X_{n+1} \in B | X_1, \dots, X_n) = \frac{\alpha_n(B)}{\alpha_n(\mathcal{X})}, \quad (4.2)$$

in cui $\alpha_n = \alpha + \sum_{i=1}^n \delta_{X_i}$, con δ_x che denota ancora una volta la misura unitaria concentrata in x . Notiamo che, nel caso di uno spazio finito \mathcal{X} , la

successione $\{X_n\}$ rappresenta il risultato di estrazioni successive da un'urna contenente inizialmente $\alpha(x)$ biglie di colore x e in cui, ad ogni estrazione, la biglia estratta viene rimessa nell'urna assieme ad un'altra biglia del medesimo colore. Osserviamo inoltre che, senza restringerci al caso \mathcal{X} finito, per ogni funzione boreliana ϕ su \mathcal{X} , la successione $\{\phi(X_n)\}$ è una successione di Pólya di parametro $\phi\alpha$, con $\phi\alpha(A) = \alpha\{\phi \in A\}$.

Passiamo allora a descrivere i collegamenti tra le successioni di Pólya e i processi di Dirichlet.

Teorema 4.1. *Sia $\{X_n\}$ una successione di Pólya di parametro μ . Allora*

- (1) $m_n = \alpha_n / \alpha_n(\mathcal{X})$ converge quasi certamente ad una misura limite discreta μ per $n \rightarrow \infty$;
- (2) μ è un processo di Dirichlet di parametro α ;
- (3) data μ , le variabili aleatorie X_1, X_2, \dots sono indipendenti con legge μ .

DIMOSTRAZIONE. Supponiamo inizialmente che \mathcal{X} sia finito,

$$\mathcal{X} = \{1, 2, \dots, r\}.$$

Siano μ e $\{X_n\}$ tali che valgano le (4.1) e (4.1). Se π_n è la distribuzione di X_1, \dots, X_n , allora dalla legge forte dei grandi numeri segue che $\pi_n \rightarrow \mu$ con probabilità uno al tendere di $n \rightarrow \infty$. Abbiamo allora la (4.1), poiché

$$m_n = \frac{\alpha + n\pi_n}{\alpha(\mathcal{X}) + n}.$$

Rimane da mostrare che $\{X_n\}$ è una successione di Pólya di parametro α , il che è equivalente a

$$\mathbb{P}(A) = \frac{\prod_x \alpha(x)^{n(x)}}{\alpha(\mathcal{X})^{[n]}}, \quad (4.3)$$

in cui $A = \{X_1 = x_1, \dots, X_n = x_n\}$, $n(x)$ indica il numero di indici i per cui $x_i = x$ e $a^{[k]} = a(a+1) \cdots (a+k-1)$.

Siccome $\mathbb{P}(A|\mu) = \prod_x \mu(x)^{n(x)}$, abbiamo

$$\mathbb{P}(A) = \mathbb{E} \left(\prod_x \mu(x)^{n(x)} \right). \quad (4.4)$$

Il fatto che i secondi termini delle (4.3) e (4.4) siano uguali è un'identità nota per i momenti delle distribuzioni di Dirichlet.

Passiamo ora al caso in cui \mathcal{X} sia generico; siano $\{X_n\}$ una successione di Pólya di parametro α ed I_j l'indicatrice dell'evento “ X_j è diverso da tutti gli X_i con $i < j$ ”. Definiamo

$$f_{nj} = \begin{cases} I_j m_n(X_j) & \text{per } 1 \leq j \leq n \\ 0 & \text{per } j > n. \end{cases}$$

Mostriamo ora che

$$f_{nj} \rightarrow f_j^* \text{ q.c. per } n \rightarrow \infty \quad (4.5)$$

per qualche f_j^* e

$$\sum_j f_j^* = 1 \quad \text{q.c.} \quad (4.6)$$

Il punto (4.1) dell'enunciato, con μ definito da

$$\mu(B) = \sum_{x_j \in B} f_j^*,$$

è un'immediata conseguenza delle (4.5) e (4.6) dal momento che, per ogni B , ponendo

$$s_{nr} = \sum_{1 \leq j \leq r; x_j \in B} f_{nj}, \quad t_{nr} = \sum_{1 \leq j \leq r} f_{nj}$$

abbiamo

$$s_{nr} \leq m_r(B) \leq s_{nr} + (1 - t_{nr}) \quad 1 \leq r \leq n, \quad (4.7)$$

cosicché mandando prima $n \rightarrow \infty$, poi $r \rightarrow \infty$ otteniamo la (4.1).

Per ottenere la (4.5) e la (4.6), fissiamo r e definiamo

$$U_n = \begin{cases} j & \text{se } 1 \leq j \leq n \text{ e } I_j = 1 \text{ e } X_{r+n} = X_j \\ 0 & \text{altrimenti.} \end{cases}$$

Dati X_1, \dots, X_r , la successione $\{U_n\}$ è di Pólya su $\{0, 1, \dots, r\}$ di parametro α' definito da

$$\begin{aligned} \alpha'(j) &= \alpha_r(\mathcal{X}) f_{rj} \quad 1 \leq j \leq r \\ \alpha'(0) &= \alpha_r(\mathcal{X}) - \sum_{j=1}^r \alpha'(j) \end{aligned}$$

e la successione m'_n associata a $\{U_n\}$ soddisfa le

$$m'_n(j) = f_{r+n, j} \quad 1 \leq j \leq r \quad (4.8)$$

e

$$m'_n(0) = 1 - \sum_{j=1}^r f_{r+n, j}. \quad (4.9)$$

Possiamo ora applicare il caso finito del teorema ad $\{U_n\}$. Dalla (4.8) e dalla (4.1) abbiamo la (4.5), mentre dalla (4.9) e dalla (4.1) otteniamo

$$\mathbb{E} \left[1 - \sum_{j=1}^r f_j^* | X_1, \dots, X_r \right] = \frac{\alpha'(0)}{\alpha'_r(\mathcal{X})} \leq \frac{\alpha(\mathcal{X})}{\alpha(\mathcal{X}) + r}. \quad (4.10)$$

Se prendiamo la speranza nella (4.10) e mandiamo $r \rightarrow \infty$, otteniamo $\mathbb{E} \left[1 - \sum_{j=1}^{\infty} f_j^* \right] = 0$, da cui segue la (4.6).

I punti (4.1) e (4.1) del teorema sono ora semplici conseguenze del caso finito. Per ogni partizione finita B_1, \dots, B_r di \mathcal{X} , definiamo ϕ su \mathcal{X} come $\phi = i$ su B_i , di modo che $\{\phi(X_n)\}$ risulta essere una successione di Pólya di parametro $\phi\alpha$. A questo punto concludiamo osservando che il limite di $(m_n(B_1), \dots, m_n(B_r))$, già identificato con $(m^*(B_1), \dots, m^*(B_r))$, ha distribuzione di Dirichlet di parametri $\alpha(B_1), \dots, \alpha(B_r)$, fatto che ci dà la (4.1). Per il punto (4.1), sia $\{\phi_i\}$ una successione di funzioni su \mathcal{X} , ciascuna delle quali assuma solo un numero finito di valori, tale che se \mathcal{F}_i è la tribù determinata da ϕ_i , abbiamo $\mathcal{F}_{i+1} \supset \mathcal{F}_i$ e la tribù determinata da $\mathcal{F} = \cup \mathcal{F}_i$ è costituita da boreliani. Allora il punto (4.1) del caso finito del teorema, applicato alla $\{\phi_j(X_n)\}$, ci dà

(3') data $\phi_j\mu$, la successione $\{\phi_i(X_n)\}$ è indipendente con legge $\phi_i\mu$ per $i \leq j$.

Mandando $j \rightarrow \infty$, otteniamo

(3'') data μ , la successione $\{\phi_i(X_n)\}$ è indipendente con legge $\phi_i\mu$ per ogni i .

Dal momento che $\{\phi_i(X_n)\}$ è indipendente con legge $\phi_i\mu$ per ogni i implica che $\{X_n\}$ è indipendente con legge μ , il punto (4.1) segue da (4.1), completando la dimostrazione. \square

Osservazione 4.1. Possiamo notare che da questa tecnica costruttiva otteniamo immediatamente, grazie alle proprietà delle successioni di Pólya, due importanti proprietà, già citate nel capitolo 3. La prima è che, estratto un campione X di dimensione 1 da un processo di Dirichlet μ di parametro α , abbiamo, per $A \in \mathcal{A}$

$$\mathbb{P}(X \in A) = \frac{\alpha(A)}{\alpha(\mathcal{X})}.$$

Inoltre se estraiamo da μ un campione di dimensione n , X_1, \dots, X_n , la distribuzione condizionale di μ dati X_1, \dots, X_n è ancora un processo di Dirichlet, di parametro

$$\alpha + \sum_{i=1}^n \delta_{X_i},$$

ossia, se estraiamo un altro campione di dimensione 1, X ,

$$\mathbb{P}(X \in A | X_1, \dots, X_n) = \frac{\alpha(A) + \sum_{i=1}^n \delta_{X_i}(A)}{\alpha(\mathcal{X}) + n}. \quad (4.11)$$

4.2 Stick-breaking

Sia α una misura finita non identicamente nulla su $(\mathcal{X}, \mathcal{B})$; sia inoltre $\tilde{\alpha}(B) = \alpha(B)/\alpha(\mathcal{X})$ la misura di probabilità normalizzata che si ottiene da α . Denotiamo con $\text{Beta}(\gamma, \delta)$ la distribuzione Beta su $[0, 1]$ di parametri γ e δ . Siano

$$\mathcal{N} = \{1, 2, \dots\}$$

e \mathcal{F} la tribù di tutti i sottoinsiemi di \mathcal{N} .

Prendiamo uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ che sia di supporto ad una famiglia di variabili aleatorie $(\vartheta, \mathbf{Y}) = ((\vartheta_i, Y_i), i \in \mathcal{N})$ con distribuzione definita nel modo seguente: le v.a. $(\vartheta_1, \vartheta_2, \dots)$ sono i.i.d. di distribuzione comune $\text{Beta}(1, \alpha(\mathcal{X}))$; le v.a. (Y_1, Y_2, \dots) sono indipendenti dalle ϑ_i e i.i.d. tra loro con distribuzione comune $\tilde{\alpha}$.

Siano ora

$$\begin{cases} p_1 = \vartheta_1 \\ p_n = \vartheta_n \prod_{m=1}^{n-1} (1 - \vartheta_m) \quad n \geq 2 \end{cases}$$

Osserviamo che l'esistenza di $(\Omega, \mathcal{A}, \mathbb{P})$ si ottiene dalla costruzione della misura prodotto e non richiede restrizioni sullo spazio $(\mathcal{X}, \mathcal{B})$, come ad esempio accade con la costruzione di Blackwell-McQueen vista nella sezione 4.1.

Definiamo allora

$$\mu_B(\vartheta, \mathbf{Y}) = \mu_B = \sum_{n=1}^{\infty} p_n \delta_{Y_n}(B) \quad (4.12)$$

in cui con $\delta_x(\cdot)$ indichiamo la misura di probabilità concentrata in x .

Questa costruzione, che vedremo essere equivalente a quella presentata nel capitolo 3, prende anche il nome di *stick-breaking*. Possiamo infatti visualizzare la costruzione dei pesi p_i nel modo seguente. Iniziando con un bastoncino di lunghezza 1, lo spezziamo in ϑ_1 e poniamo p_1 uguale alla lunghezza del primo

frammento. Ora prendiamo il secondo frammento e lo spezziamo ripetutamente in modo analogo, ottenendo così p_2, p_3 e così via. Grazie alla sua semplicità la tecnica di stick-breaking ha dato origine ad alcune generalizzazioni, come quella proposta in [FW08], nonché a nuove tecniche per i processi di Dirichlet. A questo proposito si veda anche [IJ01].

4.2.1 μ è la misura di Dirichlet

Il cammino per mostrare che quella appena definita non è altro che la misura di Dirichlet si sviluppa in alcuni passi.

Innanzitutto siano $\vartheta_n^* = \vartheta_{n+1}$, $Y_n^* = Y_{n+1}$ per $n \in \mathcal{N}$. Se poniamo $(\vartheta^*, \mathbf{Y}^*) = ((\vartheta_i^*, Y_i^*)_i)$ e

$$\begin{cases} p_1^* = \vartheta_1^* = \vartheta_2 \\ p_n^* = \vartheta_n^* \prod_{m=1}^{n-1} (1 - \vartheta_m^*) = \vartheta_{n+1} \prod_{m=2}^n (1 - \vartheta_m) \quad n \geq 2, \end{cases}$$

abbiamo che la μ definita come in (4.12) è tale che

$$\begin{aligned} \mu_B(\vartheta, \mathbf{Y}) &= p_1 \delta_{Y_1}(B) + \sum_{n=2}^{\infty} p_n \delta_{Y_n}(B) \\ &= p_1 \delta_{Y_1}(B) + (1 - \vartheta_1) \sum_{n=1}^{\infty} p_n^* \delta_{Y_n^*}(B) \\ &= \vartheta_1 \delta_{Y_1}(B) + (1 - \vartheta_1) \mu_B(\vartheta^*, \mathbf{Y}^*) \end{aligned}$$

Osserviamo che $(\vartheta^*, \mathbf{Y}^*)$ ha la stessa distribuzione di (ϑ, \mathbf{Y}) ed è indipendente da (ϑ_1, Y_1) , per come abbiamo preso le (ϑ_i, Y_i) . Possiamo allora riscrivere l'uguaglianza precedente come identità tra distribuzioni nel modo seguente:

$$\mu \stackrel{st}{=} \vartheta_1 \delta_{Y_1} + (1 - \vartheta_1) \mu \quad (4.13)$$

con la μ nel secondo membro indipendente da (ϑ_1, Y_1) .

Lemma 4.2. *Siano*

$$\gamma = (\gamma_1, \dots, \gamma_k) \quad e \quad \delta = (\delta_1, \dots, \delta_k)$$

due vettori k -dimensionali. Siano inoltre U, V due vettori aleatori indipendenti k -dimensionali aventi distribuzioni di Dirichlet $\text{Dir}(\gamma)$ e $\text{Dir}(\delta)$ rispettivamente. Sia W indipendente da (U, V) con distribuzione Beta(γ', δ'), in cui $\gamma' = \sum \gamma_i$ e $\delta' = \sum \delta_i$. Allora la distribuzione di

$$WU + (1 - W)V$$

è la distribuzione di Dirichlet $\text{Dir}(\gamma + \delta)$.

Lemma 4.3. *Siano $\gamma = (\gamma_1, \dots, \gamma_k)$, $\gamma' = \sum \gamma_i$ e $\tilde{\gamma}_i = \gamma_i/\gamma'$ per $i = 1, \dots, k$. Allora*

$$\sum \tilde{\gamma}_i \text{Dir}(\gamma + \mathbf{e}_i) = \text{Dir}(\gamma).$$

Questo fatto si può scrivere anche nel modo seguente:

$$\mathbb{E}(\text{Dir}(\gamma + \mathbf{Z})) = \text{Dir}(\gamma),$$

in cui \mathbf{Z} è il vettore aleatorio che assume i valori \mathbf{e}_i con probabilità $\tilde{\gamma}_i$.

Lemma 4.4. *Siano U e W una coppia di variabili aleatorie tali che W assume valori in $[-1, 1]$ e U in uno spazio lineare. Supponiamo che V sia una variabile aleatoria a valori nello stesso spazio lineare di U che è indipendente da (W, U) e soddisfa l'uguaglianza*

$$V \stackrel{st}{=} U + WV. \quad (4.14)$$

Se $\mathbb{Q}(|W| = 1) \neq 1$, allora esiste un'unica distribuzione per V che soddisfa la (4.14).

DIMOSTRAZIONE. Siano per assurdo V e V' due variabili aleatorie le cui leggi siano diverse, ma soddisfino entrambe la (4.14). Siano inoltre $(W_n, U_n)_n$ copie indipendenti di (W, U) che siano anche indipendenti da V e V' . Poniamo

$$\begin{cases} V_1 = V \\ V_{n+1} = U_n + W_n V_n \end{cases} \quad n \geq 1 \quad \begin{cases} V'_1 = V' \\ V'_{n+1} = U_n + W_n V'_n \end{cases} \quad n \geq 1.$$

Per come sono state definite, abbiamo che, grazie alla (4.14), tutte le V_n hanno la stessa legge di V , così come le V'_n hanno la stessa legge di V' .

Osserviamo ora che le W_n sono variabili aleatorie i.i.d., quindi

$$\mathbb{P}(|W_n| \leq 1) = \mathbb{P}(|W_1| \leq 1) = 1 \quad \mathbb{P}(|W_n| = 1) = \mathbb{P}(|W_1| = 1) < 1.$$

Allora, se consideriamo $|V_{n+1} - V'_{n+1}|$, abbiamo

$$|V_{n+1} - V'_{n+1}| = |W_n| |V_n - V'_n| = \prod_{j=1}^n |W_j| |V_1 - V'_1| \rightarrow 0.$$

Ciò contraddice l'ipotesi dell'assurdo che le leggi di V e V' fossero diverse. Pertanto abbiamo che se la legge di V soddisfa la (4.14), allora è unica. \square

Teorema 4.5. *Sia $\{B_1, \dots, B_k\}$ una partizione misurabile di \mathcal{X} e sia*

$$\mu^* = (\mu_{B_1}, \dots, \mu_{B_k}).$$

Allora la distribuzione di μ^ è la distribuzione di Dirichlet k -dimensionale $\text{Dir}(\alpha(B_1), \dots, \alpha(B_k))$.*

DIMOSTRAZIONE. Sia $D = (\delta_{Y_1}(B_1), \dots, \delta_{Y_1}(B_k))$; osserviamo che

$$\mu(D = \mathbf{e}_i) = \mu(Y_1 \in B_i) = \tilde{\alpha}(B_i) \quad i = 1, \dots, k.$$

Dalla (4.13) abbiamo che μ^* soddisfa la

$$\mu^* \stackrel{st}{=} \vartheta_1 D + (1 - \vartheta_1)\mu^* \quad (4.15)$$

in cui, nel secondo membro, ϑ_1 ha distribuzione $\text{Beta}(1, \alpha(\mathcal{X}))$, D è indipendente da ϑ_1 e assume il valore \mathbf{e}_i con probabilità $\tilde{\alpha}(B_i)$ e il vettore aleatorio k -dimensionale μ^* è indipendente da (ϑ_1, D) .

Come prima cosa verifichiamo che la distribuzione k -dimensionale di Dirichlet per μ^* soddisfa la (4.15), dopodiché mostreremo che tale soluzione è l'unica.

Supponiamo che la distribuzione di μ^* nel secondo membro della (4.15) sia la distribuzione k -dimensionale di Dirichlet $\text{Dir}(\alpha(B_1), \dots, \alpha(B_k))$. La distribuzione di Dirichlet k -dimensionale $\text{Dir}(\mathbf{e}_i)$ dà probabilità 1 ad \mathbf{e}_i . Fissato $D = \mathbf{e}_i$, la distribuzione di $\vartheta_1 D + (1 - \vartheta_1)\mu^*$ è la distribuzione di

$$\vartheta_1 \text{Dir}(\mathbf{e}_i) + (1 - \vartheta_1) \text{Dir}(\alpha(B_1), \dots, \alpha(B_k))$$

e questa, per il lemma 4.2, è $\text{Dir}((\alpha(B_1), \dots, \alpha(B_k)) + \mathbf{e}_i)$.

Sommare sulla distribuzione di D è equivalente a prendere una combinazione di queste Dirichlet con pesi $\tilde{\alpha}(B_i)$, che per il lemma 4.3 è uguale a $\text{Dir}(\alpha(B_1), \dots, \alpha(B_k))$. Questo prova che la distribuzione di Dirichlet k -dimensionale soddisfa la (4.15). Grazie al lemma 4.4 tale soluzione è unica. \square

Osservazione 4.2. A partire dalla costruzione alternativa appena data possiamo ricavare alcune importanti proprietà del processo di Dirichlet, in particolare il più volte citato fatto che μ è concentrato sulle misure discrete di probabilità.

4.3 Ristorante Cinese

Abbiamo già osservato in precedenza che, per il processo di Dirichlet, vale la proprietà che estratto da μ un campione di dimensione $n + 1$, (X_1, \dots, X_{n+1}) , allora

$$\mathbb{P}(X_{n+1} \in A | X_1, \dots, X_n) = \frac{1}{\alpha(\mathcal{X}) + n} \left(\alpha(A) + \sum_{i=1}^n \delta_{X_i}(A) \right),$$

si vedano ad esempio la (4.11) ed il teorema 3.8. Ne segue, in termini di leggi,

$$X_{n+1} | X_1, \dots, X_n \sim \frac{1}{\alpha(\mathcal{X}) + n} \left(\alpha + \sum_{i=1}^n \delta_{X_i} \right). \quad (4.16)$$

In questa sezione vogliamo mettere a fuoco questa proprietà del processo di Dirichlet per caratterizzarlo in modo alternativo, sottolineandone frattanto alcune ulteriori proprietà.

Una proprietà che segue immediatamente da quest'ultima uguaglianza è che la distribuzione del campione successivo, dati n campioni estratti, ha una componente di distribuzione atomica sui campioni precedenti. Allora con probabilità non nulla campioni diversi estratti da μ saranno uguali, indipendentemente dalla regolarità di α . Questo ci dà un indizio del fatto che μ stesso è concentrato sulle distribuzioni discrete. Ma oltre a questa, abbiamo anche una proprietà di accumulazione. Per comodità supponiamo che α sia regolare, in modo che il fatto che i campioni si ripetano sia conseguenza solamente delle proprietà del processo di Dirichlet e non del fatto che la α sia discreta. Siccome i campioni estratti si ripetono, denotiamo con X_1^*, \dots, X_m^* i campioni distinti in X_1, \dots, X_n e con n_k la molteplicità del campione X_k^* . Possiamo dunque riscrivere la (4.16) come

$$X_{n+1}|X_1, \dots, X_n \sim \frac{1}{\alpha(\mathcal{X}) + n} \left(\alpha + \sum_{k=1}^m n_k \delta_{X_k^*} \right). \quad (4.17)$$

Questo ci dice che il campione $(n+1)$ -simo X_{n+1} assumerà valore X_k^* con probabilità proporzionale ad n_k , ossia al numero di volte che tale valore è già stato assunto. Abbiamo allora la proprietà di accumulazione cercata, perché più un valore X_k^* viene assunto, più probabilmente lo sarà anche in futuro.

Osserviamo che ciò dà luogo in modo naturale ad una partizione aleatoria dell'insieme $\{1, \dots, n\}$. In effetti possiamo caratterizzare il processo di Dirichlet a partire da questa partizione. Partiamo dalla distribuzione sulle partizioni aleatorie: possiamo ricostruire la distribuzione congiunta delle X_1, \dots, X_n estraendo una partizione casuale di $\{1, \dots, n\}$, poi per ogni insieme I_k nella partizione prendiamo una $X_k^* \sim \tilde{\alpha}$ e infine poniamo $X_i = X_k^*$ per ogni $i \in I_k$. Abbiamo allora la distribuzione congiunta

$$\mathbb{P}(X_1, \dots, X_n) = \prod_{i=1}^n \mathbb{P}(X_i | X_1, \dots, X_{i-1}).$$

Ora possiamo ricostruire il processo di Dirichlet in modo assolutamente analogo a quanto fatto nella costruzione via urne di Pólya, nella sezione 4.1, osservando che le X_i sono scambiabili e invocando il teorema di De Finetti.

La distribuzione sulle partizioni prende il nome di *processo del ristorante cinese*, a causa della seguente analogia. Consideriamo un ristorante cinese con infiniti tavoli e infiniti piatti nel proprio menu. Questo ristorante sistema i suoi clienti in un modo un po' particolare: l'ultimo arrivato, cioè l' $(n+1)$ -simo cliente, viene fatto accomodare ad un tavolo già occupato con probabilità

proporzionale al numero di clienti già seduti a quel tavolo (ossia n_k), oppure ad un tavolo libero con probabilità proporzionale ad una costante ($\alpha(\mathcal{X})$). I clienti che siedono ad un tavolo già occupato ricevono il medesimo piatto che hanno gli altri clienti a quel tavolo, mentre chi siede ad un tavolo libero ottiene un piatto a scelta dello chef tra quelli non ancora proposti. Allora la distribuzione dei piatti dopo che n clienti sono stati serviti è un campione estratto come in (4.16), in cui X_i è il piatto assegnato all' i -simo cliente.

Capitolo 5

Misura entropica

In questo capitolo cambiamo radicalmente contesto: partiamo da una varietà riemanniana compatta (M, d) e su di essa consideriamo lo spazio delle misure di probabilità $\mathcal{P}(M)$. Su questo spazio possiamo definire la distanza L^2 di Wasserstein d_W ottenendo un oggetto molto interessante da un punto di vista geometrico (si vedano, ad esempio, i lavori di Brenier, McCann, Otto, Villani ed altri: [Bre91], [McC97], [CEMS01], [Ott01], [OV00], [Vil03], [Vil09]). In particolare un ruolo di primo piano è assunto dall'entropia relativa $\mathcal{P}(M) \rightarrow [0, +\infty]$ definita da

$$\text{Ent}(\mu) = \begin{cases} \int_M \rho \log \rho \, dx & \text{se } d\mu(x) \ll dx \quad \rho(x) = \frac{d\mu(x)}{dx} \\ +\infty & \text{altrimenti,} \end{cases}$$

in cui dx è la misura del volume riemanniano su M . Alcune recenti teorie (i cui contenuti esulano dagli obiettivi di questa tesi e per cui rimandiamo ai lavori di Ambrosio, von Renesse, Sturm, Jordan ed altri: [AGS08], [vRS05], [Stu06], [JKO98]) introducono un concetto di flusso gradiente nello spazio $(\mathcal{P}(M), d_W)$ e mostrano che il flusso gradiente rispetto al funzionale dell'entropia relativa costituisce una formulazione innovativa dell'equazione del calore

$$\frac{\partial}{\partial t} \mu = \Delta \mu \quad \text{su } M.$$

Più in generale il flusso gradiente di altri funzionali $S : \mathcal{P}(M) \rightarrow \mathbb{R}$ si è rivelato un ottimo strumento nello studio di equazioni differenziali alle derivate parziali di tipo dissipativo su M (si veda, al proposito, [Vil03]). Secondo alcuni recenti sviluppi (si veda [vRS08]), sembra che lo studio di una variante stocastica del moto gradiente sia collegata alla misura \mathbb{P}^β su $\mathcal{P}(M)$, detta *misura entropica*, caratterizzata formalmente da

$$d\mathbb{P}^\beta(\mu) = \frac{1}{Z} e^{-\beta \cdot \text{Ent}(\mu)} d\mathbb{P}^0(\mu), \quad (5.1)$$

per $\beta > 0$ e Z costante di normalizzazione, con \mathbb{P}^0 un'ipotetica distribuzione uniforme su $\mathcal{P}(M)$. Osserviamo che tale misura ha come potenziale proprio l'entropia definita in precedenza.

La caratterizzazione (5.1) richiama la caratterizzazione (2.5) della misura di Wiener vista nel capitolo 2. Proprio a partire da tale euristica vogliamo dare una costruzione rigorosa della misura entropica, nel caso degli spazi unidimensionali $M = [0, 1]$ e $M = S^1$, per poi suggerire una possibile generalizzazione a spazi di dimensione maggiore.

I risultati presentati in questo capitolo sono dovuti a Sturm e von Renesse ([vRS08] e [Stu09]).

5.1 Gli spazi

Per poter arrivare alla definizione della misura entropica, abbiamo bisogno di operare su diversi spazi, che andiamo ora ad introdurre.

Consideriamo uno spazio riemanniano compatto (M, d) ; in particolare trattiamo i casi unidimensionali, con $M = [0, 1]$ ed $M = S^1$, ciascuno con la rispettiva misura di Lebesgue. Vogliamo studiare lo spazio $\mathcal{P}(M)$ delle misure di probabilità su M . Iniziamo considerando lo spazio $\mathcal{P}_0 = \mathcal{P}([0, 1])$ delle misure di probabilità sull'intervallo unitario e introduciamo la distanza di Wasserstein

$$d_W(\mu, \nu) = \inf_{\gamma} \left(\iint_{[0,1]^2} |x - y|^2 \gamma(dx, dy) \right)^{1/2}, \quad (5.2)$$

in cui $\gamma \in \Gamma(\mu, \nu) \subset \mathcal{P}([0, 1]^2)$, insieme delle misure di probabilità con marginali¹ μ e ν . Lo spazio (\mathcal{P}_0, d_W) è uno spazio metrico compatto per il teorema di Prorohov.

È interessante osservare la distanza di Wasserstein nel contesto del trasporto ottimo. L'idea è che noi abbiamo due misure, che possiamo immaginare rappresentino la distribuzione della produzione di beni e della domanda. Il nostro compito è trasportare i beni dal produttore al consumatore, ossia trasformare la prima misura nella seconda. Vorremmo calcolare in qualche modo il costo del trasporto ottimale: un modo naturale è porre

$$C(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \int c(x, y) \gamma(dx, dy),$$

in cui $c(x, y)$ è il costo del trasporto di una unità di massa da x a y . Tale C non è ancora una distanza tra le misure μ e ν , per ottenere ciò dobbiamo

¹i.e. $\gamma(A \times M) = \mu(A)$ e $\gamma(M \times B) = \nu(B)$ per ogni $A, B \subset M$

prendere c come funzione della distanza nello spazio sottostante. Arriviamo così alla misura di Wasserstein sopraindicata.

Indichiamo con \mathcal{G}_0 lo spazio delle funzioni non decrescenti e continue a destra $g : [0, 1[\rightarrow [0, 1]$ e dotiamolo della distanza L^2 :

$$\|g_1 - g_2\|_{L^2} = \left(\int_0^1 |g_1(t) - g_2(t)|^2 dt \right)^{1/2}.$$

Per comodità, inoltre, estendiamo tutte le $g \in \mathcal{G}_0$ all'intero intervallo $[0, 1]$ ponendo $g(1) = 1$. La mappa

$$\chi : \mathcal{G}_0 \rightarrow \mathcal{P}_0 \quad g \mapsto g_* \text{Leb}, \quad (5.3)$$

è la misura immagine della misura di Lebesgue sull'intervallo $[0, 1]$ rispetto alla funzione g , definita da

$$(g_* \text{Leb})(B) = \text{Leb}(g^{-1}(B)) \quad B \subset [0, 1].$$

Essa è una biiezione tra \mathcal{G}_0 e \mathcal{P}_0 . La mappa inversa

$$\chi^{-1} : \mathcal{P}_0 \rightarrow \mathcal{G}_0 \quad \mu \mapsto g_\mu$$

associa ad ogni misura di probabilità $\mu \in \mathcal{P}_0$ la sua funzione di distribuzione inversa definita da

$$g_\mu(t) = \inf \{s \in [0, 1] : \mu([0, s]) > t\} \quad (5.4)$$

con $\inf \emptyset = 1$.

Per ogni $g \in \mathcal{G}_0$ possiamo definire l'*inversa destra* $g^{(-1)} \in \mathcal{G}_0$ come

$$g^{(-1)}(t) = \inf \{s \geq 0 : g(s) > t\}.$$

È interessante osservare che su \mathcal{G}_0 possiamo introdurre in modo naturale diverse topologie: la topologia L^2 su \mathcal{G}_0 visto come sottoinsieme di $L^2([0, 1], \mathbb{R})$, l'immagine della topologia debole su \mathcal{P}_0 mediante la funzione di ripartizione inversa χ^{-1} , l'immagine della topologia debole su \mathcal{P}_0 mediante la funzione di ripartizione $\mu \mapsto g_\mu^{(-1)}$. In realtà tutte queste topologie, così come molte altre, coincidono.

In modo del tutto analogo a quanto appena fatto, passiamo all'altro caso di cui abbiamo detto ci interesseremo in modo particolare: $M = S^1$. Consideriamo dunque lo spazio $\mathcal{P}_1 = \mathcal{P}(S^1)$ delle misure di probabilità sulla circonferenza unitaria: abbiamo anche in questo caso che (\mathcal{P}_1, d_W) è uno spazio metrico compatto.

Introduciamo ora lo spazio $\mathcal{G}(\mathbb{R})$ delle funzioni $g : \mathbb{R} \rightarrow \mathbb{R}$ non decrescenti e continue a destra tali che $g(x+1) = g(x) + 1$ per ogni $x \in \mathbb{R}$. Grazie a questa proprietà abbiamo che ogni funzione $g \in \mathcal{G}(\mathbb{R})$ induce in modo univoco una funzione $\pi(g) : S^1 \rightarrow S^1$. Poniamo allora $\mathcal{G} = \pi(\mathcal{G}(\mathbb{R}))$. Le funzioni di \mathcal{G} ereditano una specie di monotonicità dalle funzioni di $\mathcal{G}(\mathbb{R})$: ogni funzione $g \in \mathcal{G}$ continua mantiene l'ordine ed è omotopa alla funzione identità. Gli elementi $g \in \mathcal{G}$ prendono il nome di *funzioni monotone* di S^1 . Lo spazio \mathcal{G} è un sottospazio compatto dello spazio $L^2(S^1, S^1)$ con distanza

$$\|g_1 - g_2\|_{L^2} = \left(\int_{S^1} |g_1(t) - g_2(t)|^2 dt \right)^{1/2}.$$

In particolare è interessante lo spazio $\mathcal{G}_1 = \mathcal{G}/S^1$, in cui identifichiamo tra loro le funzioni $g, h \in \mathcal{G}$ tali che $g(\cdot) = h(\cdot + a)$ per qualche $a \in S^1$. Lo spazio \mathcal{G}_1 eredita da \mathcal{G} la metrica L^2 , che in questo caso diventa

$$\|g_1 - g_2\|_{\mathcal{G}_1} = \left(\inf_{s \in S^1} \int_{S^1} |g_1(t) - g_2(t+s)|^2 dt \right)^{1/2}.$$

Anche in questo spazio possiamo introdurre la mappa χ , definita come sopra da

$$\chi : \mathcal{G}_1 \rightarrow \mathbb{P}_1 \quad g \mapsto g_* \text{Leb}, \quad (5.5)$$

che è la misura immagine della misura di Lebesgue su S^1 rispetto alla funzione g , ottenendo una biiezione tra \mathcal{G}_1 e \mathcal{P}_1 . La mappa inversa

$$\chi^{-1} : \mathcal{P}_1 \rightarrow \mathcal{G}_1 \quad \mu \mapsto g_\mu$$

associa ad ogni misura di probabilità $\mu \in \mathcal{P}_1$ la sua funzione di distribuzione inversa definita come nella (5.4).

Arriviamo così al seguente importante risultato.

Proposizione 5.1. *Le mappe χ definite nelle (5.3) e (5.5) e le loro inverse χ^{-1} sono isometrie tra gli spazi $(\mathcal{G}_0, \|\cdot\|_{L^2})$ e (\mathcal{P}_0, d_W) e $(\mathcal{G}_1, \|\cdot\|_{\mathcal{G}_1})$ e (\mathcal{P}_1, d_W) rispettivamente. In particolare \mathcal{G}_1 è compatto.*

DIMOSTRAZIONE. Abbiamo già osservato che le χ e χ^{-1} sono biettive in entrambi i casi. Ci limitiamo al caso $M = S^1$ e proviamo che

$$d_W(\mu, \nu) = \|g_\mu - g_\nu\|_{\mathcal{G}_1} \quad \forall \mu, \nu \in \mathcal{P}_1. \quad (5.6)$$

Osserviamo che ci basta dimostrarlo nel caso di μ e ν assolutamente continue (o, equivalentemente, per g_μ e g_ν strettamente crescenti) per densità in \mathcal{P}_1 (o in \mathcal{G}_1). Sfruttiamo il punto di vista del trasporto ottimo introdotto in precedenza:

per le misure μ e ν esiste una mappa $F : S^1 \rightarrow S^1$ detta mappa di trasporto che minimizza il costo di trasporto (si veda ad esempio [Vil03] oppure [Vil09]). Fissiamo un punto in S^1 , per esempio 0, e poniamo $s = F(0)$. Allora la mappa F è una mappa di trasporto della massa μ sul segmento $]0, 1[$ sulla massa ν sul segmento $]s, s + 1[$. Dal momento che questi segmenti sono isometrici, i risultati ottenuti nel caso del segmento ci dicono che il costo minimo per un tale trasporto è dato da

$$\int_{S^1} |g_1(t) - g_2(t + s)|^2 dt.$$

Facendo variare $s \in S^1$ abbiamo la tesi. \square

5.2 Costruzione

Quello che a noi interessa, come già detto in precedenza, è introdurre una misura \mathbb{P}^β sullo spazio di Wasserstein $\mathcal{P}(M)$ caratterizzata formalmente dalla

$$d\mathbb{P}^\beta(\mu) = \frac{1}{Z} e^{-\beta \cdot \text{Ent}(\mu)} d\mathbb{P}^0(\mu), \quad (5.7)$$

per qualche (inesistente) misura uniforme \mathbb{P}^0 su $\mathcal{P}(M)$. Il nostro obiettivo è quello di dare una costruzione rigorosa di questa misura \mathbb{P}^β nel caso unidimensionale, per dare poi un cenno della costruzione nel caso multidimensionale.

L'idea è la seguente: grazie alle isometrie χ introdotte in precedenza, possiamo trasformare il problema dal costruire la misura \mathbb{P}^β su \mathcal{P}_0 (o \mathcal{P}_1) al costruire misure \mathbb{Q}_0^β (o \mathbb{Q}_1^β) su \mathcal{G}_0 (o \mathcal{G}_1). Ad esempio se consideriamo \mathcal{G}_0 , possiamo riscrivere la caratterizzazione formale (5.7) in termini della misura \mathbb{Q}_0^β come

$$\mathbb{Q}_0^\beta(dg) = \frac{1}{Z} e^{-\beta \cdot S(g)} \mathbb{Q}^0(dg), \quad (5.8)$$

in cui \mathbb{Q}^0 indica una qualche misura uniforme su $\mathcal{G}_0 \subset L^2([0, 1])$ ed $S : \mathcal{G}_0 \rightarrow [0, +\infty]$ è il funzionale dell'entropia definito come

$$S(g) = \text{Ent}(g_* \text{Leb}). \quad (5.9)$$

Osserviamo che, per le funzioni g assolutamente continue, possiamo scrivere $S(g)$ in modo esplicito come

$$S(g) = - \int_0^1 \log g'(t) dt.$$

Infatti ricordiamo che

$$\text{Ent}(g_*\text{Leb}) = \text{Ent}(\text{Leb}(g^{-1}))$$

e l'entropia relativa è definita da

$$\text{Ent}(\mu) = \begin{cases} \int_M \rho \log \rho \, dx & \text{se } d\mu(x) \ll dx \quad \rho(x) = \frac{d\mu(x)}{dx} \\ +\infty & \text{altrimenti,} \end{cases}$$

da cui segue

$$\begin{aligned} S(g) &= \int_0^1 \frac{dg^{-1}(x)}{dx} \log \left(\frac{dg^{-1}(x)}{dx} \right) dx \\ &= - \int_0^1 \frac{1}{g'(g^{-1}(x))} \log(g'(g^{-1}(x))) dx \\ &= - \int_0^1 \log(g'(t)) dt, \end{aligned}$$

con il cambio di variabile $t = g^{-1}(x)$.

La (5.8) richiama la caratterizzazione euristica della misura di Wiener, vista al capitolo 2: possiamo allora appoggiarci a quest'ultima per arrivare alla \mathbb{Q}_0^β . Abbiamo visto che la misura di Wiener \mathbb{P}^W sullo spazio \mathcal{G}^* di tutte le mappe continue $g : [0, 1] \rightarrow \mathbb{R}$ con $g(0) = 0$ può essere interpretata come

$$\mathbb{P}^W(dg) = \frac{1}{Z} e^{-H(g)} \mathbb{P}^{\mathcal{G}^*}(dg), \quad (5.10)$$

in cui $H(g)$ è il funzionale dell'energia, mentre $\mathbb{P}^{\mathcal{G}^*}(dg)$ è una "distribuzione uniforme" sullo spazio \mathcal{G}^* .

Abbiamo visto come superare l'ostacolo dell'assenza di una distribuzione siffatta mediante interpolazione; abbiamo, a partire dalla (5.10), la rappresentazione esplicita per le leggi finito dimensionali:

$$\begin{aligned} \mathbb{P}^W(g_{t_1} \in dx_1, \dots, g_{t_N} \in dx_N) &= \\ &= \frac{1}{Z_N} \exp \left(-\frac{1}{2} \sum_{i=1}^N \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}} \right) p_N(dx_1, \dots, dx_N), \end{aligned} \quad (5.11)$$

con p_N la misura N -dimensionale di Lebesgue, in modo da avere, come limite proiettivo, proprio la misura di Wiener \mathbb{P}^W su \mathcal{G}^* .

Il nostro piano è ora quello di seguire un approccio analogo sostituendo il funzionale d'entropia $S(g)$ a quello dell'energia $H(g)$. Prendiamo una qualunque partizione finita $\{0 = t_0 < \dots < t_{N+1} = 1\}$ di $[0, 1]$ ed andiamo a sostituire

l'entropia $S(g)$ della mappa g con l'entropia dell'interpolazione lineare a tratti di g :

$$\begin{aligned} S_N(g) &= \inf \{S(\tilde{g}) : \tilde{g} \in \mathcal{G}^*, \tilde{g}(t_i) = g(t_i) \forall i\} \\ &= - \sum_{i=1}^{N+1} \log \frac{|g(t_i) - g(t_{i-1})|^2}{2(t_i - t_{i-1})} \cdot (t_i - t_{i-1}). \end{aligned}$$

In questo modo abbiamo la seguente espressione delle leggi finito dimensionali:

$$\begin{aligned} &\mathbb{Q}_0^\beta(g_{t_1} \in dx_1, \dots, g_{t_N} \in dx_N) = \\ &= \frac{1}{Z_N} \exp \left(\beta \sum_{i=1}^{N+1} \log \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}} \cdot (t_i - t_{i-1}) \right) q_N(dx_1, \dots, dx_N), \end{aligned} \quad (5.12)$$

dove $q_N(dx_1, \dots, dx_N) = \mathbb{Q}^0(g_{t_1} \in dx_1, \dots, g_{t_N} \in dx_N)$ è una "distribuzione uniforme" sul simpleso

$$\Delta_N = \left\{ (x_1, \dots, x_N) \in [0, 1]^N : 0 < x_1 < \dots < x_N < 1 \right\}$$

con $x_0 = 0$ e $x_{N+1} = 1$.

Dobbiamo ora determinare una q_N adeguata. Una richiesta naturale è che soddisfi la seguente proprietà di invarianza:

$$\begin{aligned} q_N(dx_1, \dots, dx_N) &= ((R^{x_{i-1}, x_{i+k}})_* q_k(dx_i, \dots, dx_{i+k-1})) \\ &\quad dq_{N-k}(dx_1, \dots, dx_{i-1}, dx_{i+k}, \dots, dx_N), \end{aligned} \quad (5.13)$$

per ogni $k = 1, \dots, N$ ed $i = 1, \dots, N - k + 1$, con la convenzione che $x_0 = 0$ e $x_{N+1} = 1$ e in cui $R^{a,b}$ è la funzione di riscaldamento definita da

$$R^{a,b} :]0, 1[^k \rightarrow \mathbb{R}^k, \quad y_j \mapsto y_j(b - a) + a, \quad j = 1, \dots, k.$$

Se le q_N , al variare di $N \in \mathbb{N}$, fossero misure di probabilità, la proprietà d'invarianza (5.13) potrebbe essere interpretata nel modo seguente, che ne giustifica l'introduzione: sotto q_N la legge del blocco $(x_1, \dots, x_{i-1}, x_{i+k}, \dots, x_N)$ non è altri che q_{N-k} e la legge del blocco (x_i, \dots, x_{i+k-1}) di punti contenuti nell'intervallo $]x_{i-1}, x_{i+k}[$ è, dopo aver riscaldato l'intervallo, q_k . Purtroppo non esiste alcuna famiglia di misure di probabilità q_N al variare di $N \in \mathbb{N}$ per cui valga la proprietà (5.13). Tuttavia esiste una famiglia di misure per cui essa vale.

Iterando la proprietà (5.13), la misura q_N al variare di $N \in \mathbb{N}$ è determinata dalla scelta della misura q_1 sull'intervallo $\Delta_1 =]0, 1[$. Inoltre se consideriamo

la proprietà (5.13) nel caso $N = 2$ e $k = 1$, le due possibili scelte di i ci danno la seguente uguaglianza:

$$\left((R^{0,x_1})_* q_1(dx_2) \right) dq_1(dx_1) = \left((R^{x_2,1})_* q_1(dx_1) \right) dq_1(dx_2), \quad (5.14)$$

per ogni $0 < x_1 < x_2 < 1$. Questo rispecchia la ragionevole richiesta che sia equivalente scegliere prima $x_1 \in]0, 1[$ e poi $x_2 \in]x_1, 1[$ o, viceversa, prima $x_2 \in]0, 1[$ e poi $x_1 \in]0, x_2[$.

Arriviamo allora al seguente risultato, che caratterizza la famiglia di misure q_N cercata.

Lemma 5.2. *Una famiglia di misure q_N , al variare di $N \in \mathbb{N}$, con densità continue soddisfa la proprietà (5.13) se e solo se*

$$q_N(dx_1, \dots, dx_N) = C^N \frac{dx_1 \cdots dx_N}{x_1 \cdot (x_2 - x_1) \cdots (x_N - x_{N-1}) \cdot (1 - x_N)}, \quad (5.15)$$

per qualche costante $C \in \mathbb{R}_+$.

DIMOSTRAZIONE. Se $q_1(dx) = \rho(x) dx$, allora la (5.14) si può scrivere come

$$\rho(y) \cdot \rho\left(\frac{x}{y}\right) \cdot \frac{1}{y} = \rho(x) \cdot \rho\left(\frac{y-x}{1-x}\right) \cdot \frac{1}{1-x},$$

per ogni $0 < x < y < 1$. Se ρ è continua, da questa uguaglianza abbiamo che esiste una costante $C \in \mathbb{R}_+$ tale che

$$\rho(x) = \frac{C}{x(1-x)},$$

per ogni $0 < x < 1$. Se inseriamo ripetutamente questo risultato nella (5.13), otteniamo la tesi. \square

Ricordiamo che il nostro obiettivo è quello di dare un significato alla formula euristica (5.8). Se ora andiamo a combinare la (5.12) con la scelta della q_N fatta nella (5.15) otteniamo

$$\begin{aligned} \mathbb{Q}_0^\beta(g_{t_1} \in dx_1, \dots, g_{t_N} \in dx_N) &= \\ &= \frac{1}{Z_N} \prod_{i=1}^{N+1} (x_i - x_{i-1})^{\beta(t_i - t_{i-1})} \frac{dx_1 \cdots dx_N}{x_1 \cdot (x_2 - x_1) \cdots (1 - x_N)} \end{aligned} \quad (5.16)$$

con un'opportuna costante di normalizzazione Z_N . Il secondo membro della (5.16) definisce una famiglia di misure di probabilità. Inoltre, facendo qualche

manipolazione, abbiamo

$$\begin{aligned} \mathbb{Q}_0^\beta (g_{t_1} \in dx_1, \dots, g_{t_N} \in dx_N) &= \\ &= \frac{1}{Z_N} \prod_{i=1}^{N+1} (x_i - x_{i-1})^{\beta(t_i - t_{i-1}) - 1} dx_1 \cdots dx_N, \end{aligned}$$

che con

$$Z_N = \frac{\Gamma(\beta)}{\prod_{i=1}^{N+1} \Gamma(\beta(t_i - t_{i-1}))}$$

risulta essere proprio una famiglia di distribuzioni di Dirichlet sulla partizione $\{t_0 = 0, t_1, \dots, t_N, t_{N+1} = 1\}$ dell'intervallo $[0, 1]$. Allora \mathbb{Q}_0^β è proprio il processo di Dirichlet.

Dai risultati esposti nei capitoli 3 e 4, sappiamo che possiamo costruire il processo di Dirichlet su ogni spazio. Sono allora ben date le seguenti definizioni.

Definizione 5.1. La *misura entropica* \mathbb{P}_0^β su $\mathcal{P}_0 = \mathcal{P}([0, 1])$ è definita come la misura immagine del processo di Dirichlet \mathbb{Q}_0^β di parametro $\alpha = \beta \cdot \text{Leb}$ su \mathcal{G}_0 rispetto alla mappa χ definita dalla 5.3. Ciò significa che, per tutte le funzioni misurabili limitate $u : \mathcal{P}_0 \rightarrow \mathbb{R}$,

$$\int_{\mathcal{P}_0} u(\mu) d\mathbb{P}_0^\beta(\mu) = \int_{\mathcal{G}_0} u(g_* \text{Leb}) d\mathbb{Q}_0^\beta(g).$$

Definizione 5.2. La *misura entropica* \mathbb{P}_1^β su $\mathcal{P}_1 = \mathcal{P}(S^1)$ è definita come la misura immagine del processo di Dirichlet \mathbb{Q}_1^β di parametro $\alpha = \beta \cdot \text{Leb}$ su \mathcal{G} rispetto alla mappa χ definita dalla 5.5. Ciò significa che, per tutte le funzioni misurabili limitate $u : \mathcal{P}_1 \rightarrow \mathbb{R}$,

$$\int_{\mathcal{P}_1} u(\mu) d\mathbb{P}_1^\beta(\mu) = \int_{\mathcal{G}} u(g_* \text{Leb}) d\mathbb{Q}_1^\beta(g).$$

5.3 Ulteriori considerazioni

La prima cosa che possiamo osservare è che per $M = [0, 1]$ abbiamo usato le seguenti biiezioni:

$$\mu \xrightleftharpoons{f(x)=\mu([0,x])} f \xleftarrow{g=f^{(-1)}} g,$$

partendo da una misura di probabilità, proseguendo con la sua funzione di ripartizione e arrivando alla funzione di ripartizione inversa. In realtà potremmo anche allungare la catena, nel modo seguente:

$$\mu \xrightleftharpoons{f(x)=\mu([0,x])} f \xleftarrow{g=f^{(-1)}} g \xrightleftharpoons{g(x)=\nu([0,x])} \nu,$$

ottenendo così una biiezione $\mathcal{P}_0 \rightarrow \mathcal{P}_0$. Possiamo anche riscrivere la catena in modo leggermente diverso, per evidenziare le funzioni χ che abbiamo usate:

$$\mu \xleftrightarrow{g=\chi^{-1}(\mu)} g \xleftrightarrow{g=f^{(-1)}} f \xleftrightarrow{\nu=\chi(f)} \nu.$$

Questo, unito al fatto che sappiamo costruire un processo di Dirichlet anche direttamente sullo spazio \mathcal{P}_0 , ci suggerisce una via leggermente diversa, anche se equivalente alla precedente, di definire la misura entropica. Sia infatti $\mathfrak{C} : \mathcal{P}_0 \rightarrow \mathcal{P}_0$ la mappa $\mu \mapsto \nu$: possiamo allora definire la misura entropica come la misura immagine rispetto alla mappa \mathfrak{C} del processo di Dirichlet $\widetilde{\mathbb{Q}}_0^\beta$ definito su \mathcal{P}_0 , ossia

$$\mathbb{P}_0^\beta = \mathfrak{C}_* \widetilde{\mathbb{Q}}_0^\beta.$$

Lo stesso fatto si può vedere anche nel caso $M = S^1$, in modo del tutto analogo. Tuttavia a noi interesserebbe una generalizzazione al caso M spazio compatto riemanniano di dimensione maggiore di 1. L'ostacolo ad una tale generalizzazione è costituito dalle funzioni di ripartizione e di ripartizione inversa che abbiamo adoperato in dimensione 1, poiché esse non esistono in dimensione maggiore, nonché dall'inversa destra.

Il suggerimento giusto per superare quest'impasse ci viene dalla dimostrazione della proposizione 5.1: dobbiamo guardare g non come funzione di ripartizione inversa, quanto come quell'unica mappa di trasporto ottimo tale che, nel caso unidimensionale, $\mu = g_* \text{Leb}$. Allora, in un generico spazio compatto riemanniano (M, m) in cui m sia la distribuzione uniforme normalizzata su M , potremo prendere come g quell'unica mappa di trasporto ottimo per cui $\mu = g_* m$, detta *mappa di Brenier* (si veda [Bre91]). Inoltre anche in spazi di dimensione maggiore di 1 una funzione "monotona" g siffatta ammette un'unica inversa generalizzata f , che è a sua volta una mappa "monotona" (di inversa generalizzata g). In questo modo possiamo arrivare a definire una *mappa di coniugio*

$$\mathfrak{C} : \mathcal{P}(M) \rightarrow \mathcal{P}(M), \quad \mu \mapsto \nu$$

per ogni varietà compatta M .

Arriviamo così alla definizione di misura entropica nel caso generico.

Definizione 5.3. Dati un qualunque spazio compatto riemanniano (M, m) ed un generico parametro $\beta > 0$, la *misura entropica*

$$\mathbb{P}^\beta = \mathfrak{C}_* \mathbb{Q}^\beta$$

è la misura immagine del processo di Dirichlet \mathbb{Q}^β su $\mathcal{P}(M)$ di parametro $\beta \cdot m$ rispetto alla mappa di coniugio \mathfrak{C} .

Notiamo che sia \mathbb{P}^β che \mathbb{Q}^β sono misure di probabilità sullo spazio compatto $\mathcal{P} = \mathcal{P}(M)$ delle misure di probabilità su M .

Capitolo 6

Campi Flegrei

Il processo di Dirichlet ha una certa importanza anche nella modellizzazione statistica. Infatti le sue distribuzioni finito dimensionali sono distribuzioni di Dirichlet; inoltre, come abbiamo visto, è una misura aleatoria di probabilità la cui distribuzione posteriore è a sua volta un processo di Dirichlet.

Possiamo vederne una particolare applicazione nel campo della vulcanologia. Consideriamo la regione denominata *Campi Flegrei* (d'ora in poi indicata anche con CF). Nei millenni si sono verificati vari eventi vulcanici in tale regione. Ad ogni evento eruttivo è associato un distinto punto di fuoriuscita del magma, detto *vent*. Vorremmo, a partire dai dati sperimentali raccolti, cioè dalle coordinate spaziali delle passate eruzioni, determinare una distribuzione di probabilità per il luogo del vent successivo.

6.1 Modello a due zone

Partiamo da un caso molto semplice: dividiamo la regione CF in due zone, dette CF Ovest e CF Est. Vorremmo determinare la probabilità p dell'evento "il prossimo vent è in CF Ovest"; osserviamo che, ovviamente, la probabilità dell'evento "il prossimo vent è in CF Est" è $1 - p$. Scegliamo di affrontare il problema con un approccio bayesiano. Scegliamo allora una distribuzione a priori per il valore di p ; nel nostro caso la scegliamo di classe beta, secondo un uso comune in statistica bayesiana. Allora la densità della distribuzione a priori di p è

$$f_0(p) = \frac{p^{\alpha_1^{(0)}-1} (1-p)^{\alpha_2^{(0)}-1}}{\int_0^1 t^{\alpha_1^{(0)}-1} (1-t)^{\alpha_2^{(0)}-1} dt}.$$

La scelta dei parametri $\alpha_1^{(0)}$ ed $\alpha_2^{(0)}$ è soggettiva: essa è legata alle assunzioni a priori che facciamo, come richieste di simmetria o di semplicità per il modello, oppure previsione di un certo tipo di andamento.

Fissata la distribuzione a priori entrano in gioco i dati sperimentali, in questo caso un evento vulcanico osservato in passato. Sappiamo che un certo numero di anni fa c'è stata un'eruzione, con vent diciamo in CF Ovest. Aggiorniamo allora la densità a priori f_0 in una *a posteriori* f_1 , usando la formula di Bayes:

$$f_1(p) = \frac{\mathbb{P}(\text{CFOvest}|p) f_0(p)}{\mathbb{P}(\text{CFOvest})} = \frac{p f_0(p)}{\int_0^1 p f_0(p) dp}.$$

Osserviamo che $f_1(p)$ è ancora la densità di una distribuzione di classe beta:

$$\begin{aligned} f_1(p) &= \frac{p f_0(p)}{\int_0^1 p f_0(p) dp} \\ &= p \frac{p^{\alpha_1^{(0)}-1} (1-p)^{\alpha_2^{(0)}-1}}{\int_0^1 t^{\alpha_1^{(0)}-1} (1-t)^{\alpha_2^{(0)}-1} dt} \cdot \frac{\int_0^1 t^{\alpha_1^{(0)}-1} (1-t)^{\alpha_2^{(0)}-1} dt}{\int_0^1 u \cdot u^{\alpha_1^{(0)}-1} (1-u)^{\alpha_2^{(0)}-1} du} \\ &= \frac{p^{\alpha_1^{(0)}} (1-p)^{\alpha_2^{(0)}-1}}{\int_0^1 u^{\alpha_1^{(0)}} (1-u)^{\alpha_2^{(0)}-1} du}, \end{aligned}$$

con parametri $\alpha_1^{(1)} = \alpha_1^{(0)} + 1$ e $\alpha_2^{(1)} = \alpha_2^{(0)}$.

Ovviamente, se il vent fosse stato in CF Est, avremmo avuto

$$f_1(p) = \frac{\mathbb{P}(\text{CFEst}|p) f_0(p)}{\mathbb{P}(\text{CFEst})} = \frac{(1-p) f_0(p)}{\int_0^1 (1-p) f_0(p) dp},$$

da cui ricaviamo, analogamente a prima,

$$\begin{aligned} f_1(p) &= \frac{(1-p) f_0(p)}{\int_0^1 (1-p) f_0(p) dp} \\ &= (1-p) \frac{p^{\alpha_1^{(0)}-1} (1-p)^{\alpha_2^{(0)}-1}}{\int_0^1 t^{\alpha_1^{(0)}-1} (1-t)^{\alpha_2^{(0)}-1} dt} \cdot \frac{\int_0^1 t^{\alpha_1^{(0)}-1} (1-t)^{\alpha_2^{(0)}-1} dt}{\int_0^1 (1-u) \cdot u^{\alpha_1^{(0)}-1} (1-u)^{\alpha_2^{(0)}-1} du} \\ &= \frac{p^{\alpha_1^{(0)}-1} (1-p)^{\alpha_2^{(0)}}}{\int_0^1 u^{\alpha_1^{(0)}-1} (1-u)^{\alpha_2^{(0)}} du}, \end{aligned}$$

cioè una beta di parametri $\alpha_1^{(1)} = \alpha_1^{(0)}$ e $\alpha_2^{(1)} = \alpha_2^{(0)} + 1$.

In realtà, come dati sperimentali, non abbiamo un solo evento vulcanico, bensì N : dobbiamo quindi aggiornare la densità con ciascuno di questi.

Sfruttando iterativamente le considerazioni appena fatte, otteniamo una densità a posteriori $f_N(p)$ che è ancora la densità di una distribuzione beta i cui parametri $\alpha_1^{(N)}$ ed $\alpha_2^{(N)}$ sono determinati solamente dal numero di vent in ciascuna delle due zone CF Este e CF Ovest. Se ne abbiamo registrati m in CF Ovest ed $N - m$ in CF Est, avremo una distribuzione Beta $(\alpha_1^{(N)}, \alpha_2^{(N)})$ con $\alpha_1^{(N)} = \alpha_1^{(0)} + m$ e $\alpha_2^{(N)} = \alpha_2^{(0)} + (N - m)$.

6.2 Modello a $d + 1$ zone (d -dimensionale)

Il modello unidimensionale introdotto nella sezione precedente ci mostra la strada da seguire, ma è estremamente insoddisfacente dal punto di vista applicativo: vorremmo avere una partizione un po' più dettagliata. Dividiamo allora i Campi Flegrei in $d + 1$ zone, denominate CF_1, \dots, CF_{d+1} . Vogliamo trovare il vettore $(p_1, \dots, p_d) \in [0, 1]^d$ con $\sum_{i=1}^d p_i \leq 1$ e $p_{d+1} = 1 - \sum_{i=1}^d p_i$ in cui p_i rappresenta la probabilità che il prossimo vent sia in CF_i .

Sempre seguendo l'approccio bayesiano, scegliamo come distribuzione a priori una Dirichlet, naturale generalizzazione della beta usata nel caso unidimensionale, di parametri assegnati $\alpha_1^{(0)}, \dots, \alpha_{d+1}^{(0)}$. Indichiamo con $f_0(p_1, \dots, p_d)$ la sua densità. Preso un dato sperimentale, cioè la posizione di un vent del passato, correggiamo la densità nel modo seguente: se esso è avvenuto in CF_i abbiamo

$$\begin{aligned} f_1(p_1, \dots, p_d) &= \frac{\mathbb{P}(CF_i | p_1, \dots, p_d) f_0(p_1, \dots, p_d)}{\mathbb{P}(CF_i)} \\ &= \frac{p_i f_0(p_1, \dots, p_d)}{\int p_i f_0(p_1, \dots, p_d) dp_1 \dots dp_d}. \end{aligned}$$

Procedendo in modo assolutamente analogo al caso unidimensionale (cioè esplicitando e semplificando gli integrali) otteniamo facilmente che la densità a posteriori è a sua volta una Dirichlet, di parametri

$$\left(\alpha_1^{(1)}, \dots, \alpha_{d+1}^{(1)} \right) = \left(\alpha_1^{(0)}, \dots, \alpha_i^{(0)} + 1, \alpha_{i+1}^{(0)}, \dots, \alpha_{d+1}^{(0)} \right).$$

Nel caso abbiamo a nostra disposizione N rilevazioni sperimentali, con $N = \sum_{j=1}^{d+1} m_j$ e gli m_j che indicano il numero di vent nella zona j -sima, avremo, sempre analogamente al caso unidimensionale, una distribuzione di Dirichlet di parametri $\alpha_i^{(N)} = \alpha_i^{(0)} + m_i$.

6.3 Modello continuo

Il modello d -dimensionale non è ancora completamente soddisfacente: infatti richiede che la partizione sia assegnata a priori; se si cambia la partizione bisogna ricalcolare tutto. Passiamo allora al modello continuo, basato sul processo di Dirichlet. Ora non suddividiamo più i Campi Flegrei in zone, cerchiamo invece una densità di probabilità $p(x, y)$ dove il punto (x, y) varia in CF, che possiamo vedere come un dominio bidimensionale.

Secondo l'approccio bayesiano, supponiamo di avere un processo di Dirichlet a priori, ovvero una misura finita α_0 assegnata su CF, che per semplicità supponiamo definita rispetto alla misura di Lebesgue su CF da una funzione non negativa integrabile $\alpha_0(x, y)$. Vedremo poi come scegliere tale α_0 .

Preso in esame un evento eruttivo, cioè la posizione (x_1, y_1) del suo vent, calcoliamo il processo di Dirichlet a posteriori, ovvero la funzione $\alpha_1(x, y)$. Ci interesserà poi proseguire fino ad avere una funzione associata alle N rilevazioni sperimentali in nostro possesso.

Ricordiamo come abbiamo costruito il processo di Dirichlet a partire dalle distribuzioni finito-dimensionali associate alle partizioni: queste ultime sono distribuzioni di Dirichlet. Per quanto detto nella sezione 6.2 a proposito dei modelli finito dimensionali, se il punto di venting $(x_1, y_1) \in B_i$ per un generico elemento B_i di una generica partizione, allora deve essere $\alpha_1(B_i) = \alpha_0(B_i) + 1$, mentre $\alpha_1(B_j) = \alpha_0(B_j)$ per ogni elemento B_j della partizione distinto da B_i . Quindi dobbiamo aggiungere ad $\alpha_0(x, y)$ una funzione che valga 1 se $(x_1, y_1) \in B_i$ e 0 altrimenti, qualunque siano la partizione e l'elemento della partizione. Tale funzione deve allora essere la delta di Dirac concentrata in (x_1, y_1) . Questa modellizzazione ci dà

$$\alpha_1(\cdot) = \alpha_0(\cdot) + \delta_{(x_1, y_1)}(\cdot);$$

la generalizzazione al caso in cui siano noti N eventi eruttivi diventa, di conseguenza,

$$\alpha_N(\cdot) = \alpha_0(\cdot) + \sum_{i=1}^N \delta_{(x_i, y_i)}(\cdot).$$

Osserviamo che questo modello continuo, per le proprietà del processo di Dirichlet viste, assegna ad ogni partizione una distribuzione di Dirichlet. Inoltre associa ad ogni elemento di una data partizione una distribuzione marginale beta, che misura l'incertezza della probabilità attribuita a tale regione dalla distribuzione di Dirichlet.

Come possiamo scegliere la misura α_0 di partenza, ossia la sua densità rispetto alla misura di Lebesgue? L'idea più semplice è prendere α costante su

tutto il dominio CF. In alternativa si potrebbe chiedere ad uno o più esperti di stimare a priori la distribuzione, anche assegnando α_0 costante a tratti.

6.4 Il modello in azione

Vediamo ora un'illustrazione del modello. In questo caso abbiamo scelto α_0 uniforme su tutti i Campi Flegrei. Abbiamo poi aggiornato tale misura a priori con i dati sperimentali in nostro possesso, ricavati da [ODVI04]. A questo punto abbiamo scelto una partizione della regione, mostrata nella figura 6.1. Nella zona (1.) si sono verificati 2 vent, così come nella zona (2.). Nella zona (3.) ne sono stati registrati 11, nella (4.) 21, nella (5.) 5. Infine nelle zone (6.) e (7.) ci sono stati 11 e 3 vent, rispettivamente.

A partire da questa partizione abbiamo estratto dal processo di Dirichlet la distribuzione di Dirichlet corrispondente.

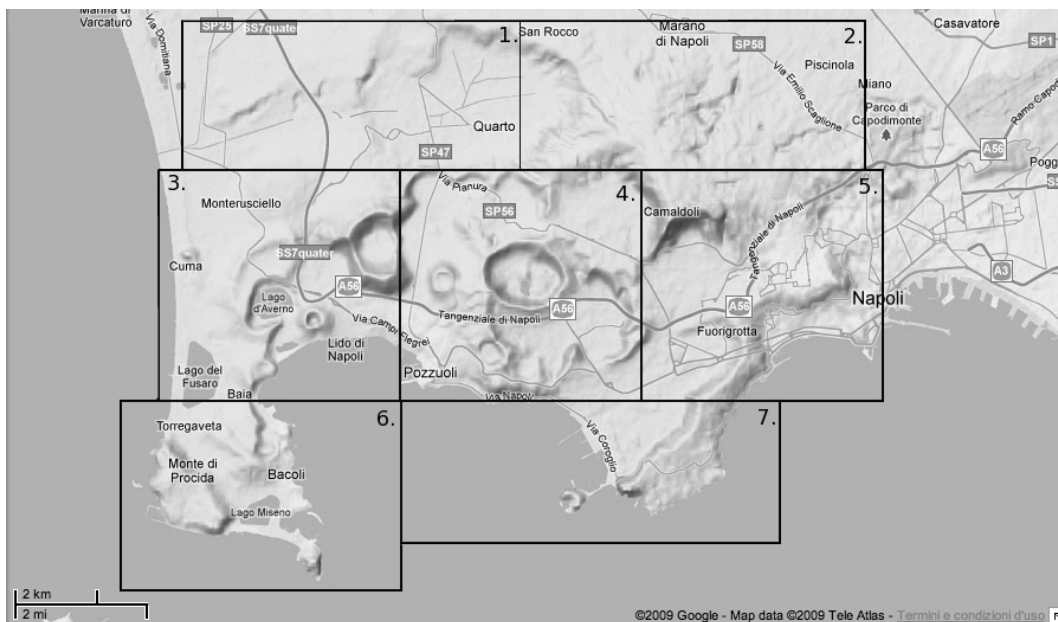
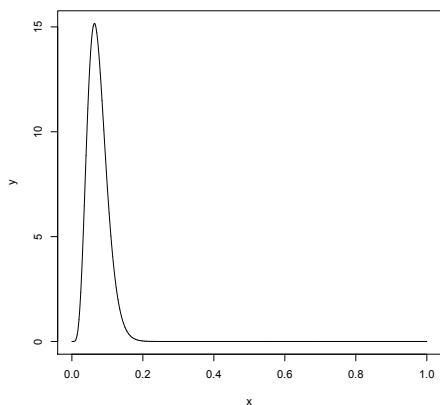


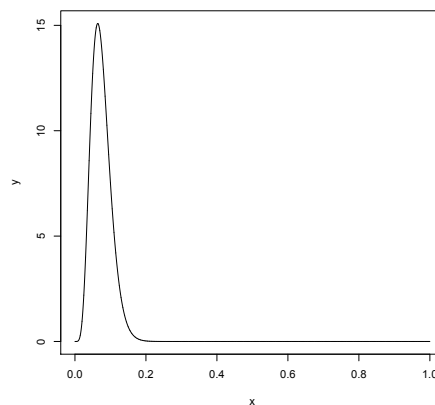
Figura 6.1: Campi Flegrei: la partizione considerata. (Immagine tratta da Google Maps, modificata dall'autore)

Nelle figure 6.2 e 6.3 sono illustrate le probabilità associate a ciascuna zona, nonché le marginali, che ci danno una misura dell'incertezza. Osserviamo che esse, per le proprietà delle distribuzioni di Dirichlet, sono distribuzioni beta. Possiamo anche notare che la probabilità che il prossimo vent si verifichi in una

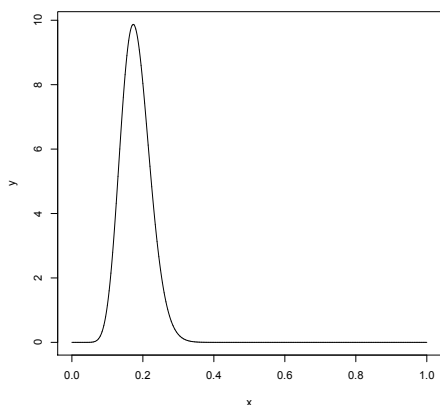
delle zone centrali (3., 4. o 6.) è sensibilmente più alta rispetto alla probabilità che esso avvenga in una delle altre zone, in accordo con il numero di eventi vulcanici già registrati in passato nelle medesime zone.



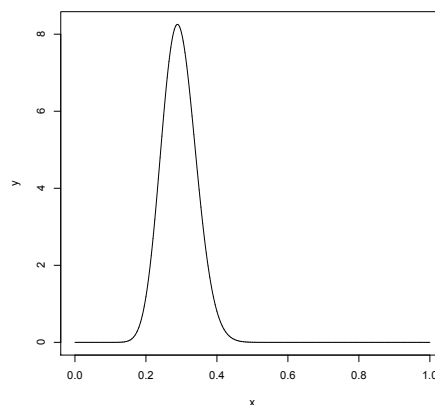
(a) Zona 1. $p = 0.0734$



(b) Zona 2. $p = 0.0744$



(c) Zona 3. $p = 0.1803$



(d) Zona 4. $p = 0.2935$

Figura 6.2: Zone - 1

6.5 Conclusioni

Sorge spontaneo chiedersi se sia vantaggioso imbarcarsi nella costruzione di un processo di Dirichlet. Un'osservazione che possiamo fare è che con questo

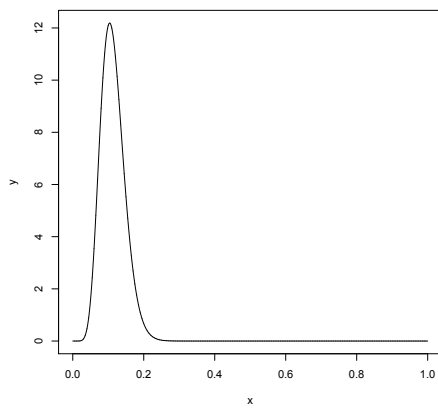
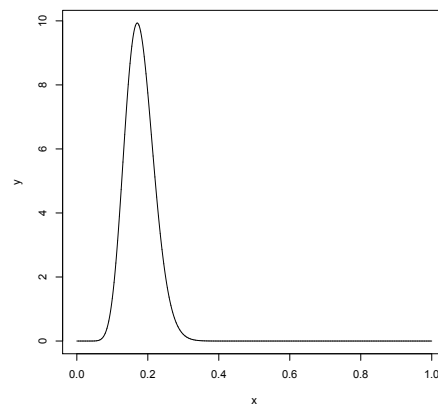
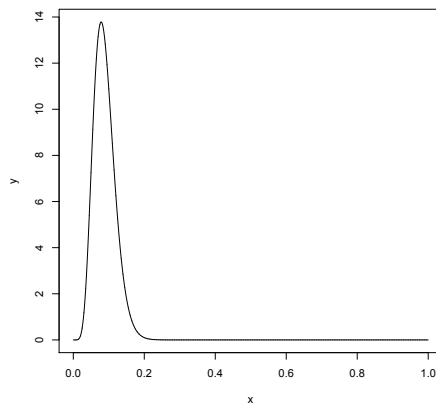
(a) Zona 5. $p = 0.1129$ (b) Zona 6. $p = 0.1777$ (c) Zona 7. $p = 0.0881$

Figura 6.3: Zone - 2

modello ad ogni zona dei Campi Flegrei è associata non solo una probabilità, ma anche una descrizione probabilistica dell'incertezza. Tale descrizione, inoltre, è ottenuta globalmente: il risultato per ogni zona tiene conto di tutte le regioni contemporaneamente. Questo è sicuramente un risultato migliore rispetto all'avere solo una descrizione locale dell'incertezza, definita zona per zona.

Si potrebbe obiettare che ciò sarebbe vero anche prendendo a priori una partizione della regione e definendo su di essa una distribuzione di Dirichlet. Un approccio di questo genere richiederebbe, però, di determinare ogni volta dei parametri opportuni per la Dirichlet rispetto alla partizione scelta, mentre nel caso del processo di Dirichlet essi si ottengono immediatamente dalla misura α , determinata una volta per tutte.

Il modello qui presentato è solamente nella fase iniziale dello sviluppo e presenta ancora molte incertezze. Ad esempio, come accennato, si potrebbe introdurre, al posto della misura α uniforme, una misura ponderata, ottenuta dalle opinioni di esperti del settore. Il rischio di una tale scelta, tuttavia, è che essa potrebbe risultare ridondante, dal momento che terrebbe intrinsecamente conto anche degli eventi vulcanici passati, che sono noti agli esperti.

Bibliografia

- [AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [Bla73] David Blackwell. Discreteness of Ferguson selections. *Ann. Statist.*, 1:356–358, 1973.
- [BM73] David Blackwell and James B. MacQueen. Ferguson distributions via Pólya urn schemes. *Ann. Statist.*, 1:353–355, 1973.
- [Bre91] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.
- [CEMS01] Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2):219–257, 2001.
- [Fer73] Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230, 1973.
- [FW08] S. Favaro and S. G. Walker. A generalized constructive definition for the Dirichlet process. *Statist. Probab. Lett.*, 78(16):2836–2838, 2008.
- [IJ01] Hemant Ishwaran and Lancelot F. James. Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.*, 96(453):161–173, 2001.
- [JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17 (electronic), 1998.

- [Kal75] Olav Kallenberg. *Random measures*. Akademie-Verlag, Berlin, 1975. Schriftenreihe des Zentralinstituts für Mathematik und Mechanik bei der Akademie der Wissenschaften der DDR, Heft 23.
- [McC97] Robert J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.
- [ODVI04] G. Orsi, M. A. Di Vito, and R. Isaia. Volcanic hazard assessment at the restless campi flegrei caldera. *Bulletin Volcanology*, 66:514–530, 2004.
- [Øks98] Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, fifth edition, 1998. An introduction with applications.
- [Ott01] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
- [OV00] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.
- [Set94] Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statist. Sinica*, 4(2):639–650, 1994.
- [Stu06] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.
- [Stu09] Karl-Theodore Sturm. Entropic measure on multidimensional spaces. Preprint on ArXiv.org, 2009.
- [Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [vRS05] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

-
- [vRS08] M.K. von Renesse and Karl-Theodore Sturm. Entropic measure and Wasserstein diffusion. To appear in *Ann. Probab.*, 2008.