ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Matematica

ESISTENZA DELLA VERSIONE REGOLARE DELLA PROBABILITÀ CONDIZIONATA

Tesi di Laurea in Probabilità e Statistica

Relatore:
Chiar.mo Prof.
ANDREA PASCUCCI

Presentata da:
ANDREA MAGNAVACCHI

II Sessione
Anno Accademico 2014/2015

A tutti coloro che meritano un ringraziamento...

Introduzione

Questa tesi nasce per trattare l'esistenza della versione regolare della probabilità condizionata ad una σ-algebra. Molti libri di testo non si soffermano su questo problema, ma si limitano ad enunciare teoremi di esistenza senza dimostrarli. Cercheremo quindi di costruire le basi per riuscire a dimostrare che in spazi usuali esiste sempre.

Nel primo capitolo partiremo definendo la misura di probabilità non direttamente su una σ-algebra, ma sull'algebra degli eventi; grazie ad alcuni risultati riusciremo ad estenderla in modo unico alla σ-algebra generata dall'algebra. Fatto ciò faremo una breve introduzione alla teoria della misura per riuscire a definire un integrale rispetto ad una misura qualunque. Per terminare introdurremo i concetti base della teoria della probabilità, come variabili aleatorie e valore atteso, probabilità condizionata a un evento e indipendenza.

Il capitolo due è incentrato invece sull'attesa condizionata rispetto ad una σ-algebra. Dimostreremo alcuni risultati preliminari come la decomposizione di Hanh per riuscire a dimostrare un importante teorema di Radon-Nikodym, grazie al quale proveremo l'esistenza dell'attesa condizionata rispetto ad una σ-algebra. Ci soffermeremo quindi sullo studio delle sue proprietà per poi evidenziarne un'importante interpretazione: potrà essere vista come proiezione negli spazi L^p, in particolare proiezione ortogonale in L^2.

Grazie all'attesa condizionata definiremo la probabilità condizionata rispetto ad una σ-algebra. Introdurremo quindi la problematica che ci porterà a definire una sua versione regolare. Esisterà tale versione?

Si può sperare che in spazi polacchi (metrici, completi e separabili) esista. Vedremo effettivamente che la risposta sarà affermativa, ma per dimostrarlo ricorreremo a un caso più generale, che ci permetterà di verificarne l'esistenza anche in altri spazi.

Indice

Capitolo 1

Cenni di teoria della misura e della probabilità

1.1 Probabilità su algebre e teorema di estensione

Definizione 1.1. (Algebra)
Sia Ω un insieme e sia $\mathcal{A}$ una famiglia di sottoinsiemi di Ω. Diremo che $\mathcal{A}$ è un'algebra se

1. $\Omega \in \mathcal{A}$
2. Se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$
3. Se $A, B \in \mathcal{A}$ allora $A \cup B \in \mathcal{A}$

Osservazione 1. Dalla definizione segue che un'algebra è chiusa rispetto all'unione finita, che $\emptyset \in \mathcal{A}$ e che è chiusa anche rispetto all'intersezione finita, infatti $A \cap B = (A^c \cup B^c)^c$

Definizione 1.2. (Probabilità su un'algebra)
Una probabilità P su un'algebra $\mathcal{A}$ di sottoinsiemi di Ω è una mappa da $\mathcal{A}$ a $[0, 1]$ tale che:

1. $0 \leq P(A) \leq 1$ per ogni $A \in \mathcal{A}$ e $P(\Omega) = 1$;

2. $P(\bigcup_I A_i) = \sum_I P(A_i)$ per ogni famiglia finita $(A_i)_I$ con $A_i \cap A_j = \emptyset$ se $i \neq j$

3. per ogni successione $(A_n)_{n\geq 1}$ decrescente a $\emptyset$, cioè tale che $A_{n+1} \subset A_n$ e $\bigcap_n A_n = \emptyset$, si ha $\lim_{n\to\infty} P(A_n) = 0$

Osservazione 2. Dalla definizione segue che $P(\emptyset) = 0$. Inoltre se $A_1 \subset A_2$ allora $P(A_1) \leq P(A_2)$, infatti $A_2 = A_1 \cup (A_2 \setminus A_1)$. Dalle decomposizioni

$$A_1 = (A_1 \cap A_2) \cup (A_1 \cap A_2^c),$$

$$A_2 = (A_2 \cap A_1) \cup (A_2 \cap A_1^c)$$

e

$$A_1 \cup A_2 = (A_1 \cap A_2) \cup (A_1 \cap A_2^c) \cup (A_1^c \cap A_2)$$

segue che $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

Lemma 1.1.1. *Se $A_n \searrow A$ in $\mathcal{A}$, allora $P(A_n) \searrow P(A)$. Se $A_n \nearrow A$ in $\mathcal{A}$, allora $P(A_n) \nearrow P(A)$.*

Dimostrazione. $A_n = (A_n \setminus A) \cup A$ e $(A \setminus A_n) \searrow \emptyset$ in $\mathcal{A}$ se $A_n \searrow A$. Per l'additività e poiché $P(A \setminus A_n) \searrow 0$ ho la prima affermazione. La seconda si dimostra dello stesso identico modo. □

Proposizione 1.1.2. *Condizione necessaria e sufficiente affinché un'applicazione P da $\mathcal{A}$ a $[0,1]$ sia una probabilità è che*

1. *$0 \leq P(A) \leq 1$ per ogni $A \in \mathcal{A}$ e $P(\Omega) = 1$;*

2. *P sia σ-additiva, cioè $P(\bigcup_I A_i) = \sum_I P(A_i)$ per ogni famiglia numerabile $(A_i)_I$ con $A_i \cap A_j = \emptyset$ se $i \neq j$ tale che $\bigcup_I A_i \in \mathcal{A}$*

Dimostrazione. Partiamo facendo vedere che è necessaria. Per il lemma precedente e per l'additività si ha che

$$P(\bigcup_{i\geq 1} A_i) = \lim_{n\to\infty} P(\bigcup_{i=1}^{n} A_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i\geq 1} P(A_i).$$

Ora vediamo che è anche sufficiente. Consideriamo $(A_n)_{n\geq 1}$ una successione decrescente a $\emptyset$. Poiché $A_n = \bigcup_{m\geq n}(A_m \setminus A_{m+1})$ e $P(A_n) \leq 1$ abbiamo

$$P(A_n) = \sum_{m\geq n} P(A_m \setminus A_{m+1}) \searrow 0.$$

□

Se P è una misura di probabilità allora $(\Omega, \mathcal{A}, P)$ è uno spazio di probabilità.

Un elemento $A \in \mathcal{A}$ viene chiamato evento e $P(A)$ è detta probabilità dell'evento A.

Si dice che un evento A è trascurabile se $P(A) = 0$. Un evento è certo se la sua probabilità vale 1.

Dopo questo studio elementare della probabilità definita su un'algebra, cercheremo di estendere la probabilità a una classe chiusa rispetto all'unione infinita. Diamo quindi la definizione di σ-algebra.

Definizione 1.3. (σ-algebra)
Una σ-algebra $\mathcal{F}$ è una famiglia di sottoinsiemi di Ω tale che

1. $\emptyset \in \mathcal{F}$
2. se $G \in \mathcal{F}$ allora $G^c \in \mathcal{F}$
3. per ogni successione $(G_n)_{n\in\mathbb{N}}$ di elementi di $\mathcal{F}$, si ha $\bigcup_n G_n \in \mathcal{F}$.

Osservazione 3. Notiamo subito che l'intersezione di due σ-algebre è ancora una σ-algebra. Chiamiamo allora, data una famiglia M di sottoinsiemi di Ω, $\sigma(M)$ la chiamiamo σ-algebra generata da M, cioè l'intersezione di tutte le σ-algebre contenti M. Questa è la più piccola sigma algebra contenente M

Esempio 1.1. *(σ-algebra di Borel)*
Sia $d \in \mathbb{N}$. La σ-algebra $\mathcal{B} = \mathcal{B}(\mathbb{R}^d)$ generata dalla topologia Euclidea di $\mathbb{R}^d$ è detta σ-algebra di Borel in $\mathbb{R}^d$. Questa è la più piccola σ-algebra contenente tutti gli aperti di $\mathbb{R}^d$. In realtà possiamo dare una definizione più generale estendendola ad un qualunque insieme E con una topologia T. Si dice che

$\mathcal{B}(E)$ è la σ-algebra generata dalla topologia T, cioè è la più piccola σ-algebra contenente tutti gli aperti di E.

Torniamo ora alla nostra probabilità definita su un'algebra $\mathcal{A}$ di Ω. Definiamo $\mathcal{G}$ come la classe di tutte le unioni delle famiglie numerabili di sottoinsiemi di Ω in $\mathcal{A}$. Se P è una probabilità su $(\Omega, \mathcal{A})$ possiamo definire una funzione Q su $\mathcal{G}$ attraverso $Q(G) := \lim_{n\to\infty} P(A_n)$ se $A_n \to G$ con $A_n \in \mathcal{A}$.

In realtà questo può essere fatto solo grazie al seguente lemma che ci assicura che il limite non dipende dalla successione che tende a G.

Lemma 1.1.3. *Se $(A_n)_{n\geq 1}$ e $(A'_m)_{m\geq 1}$ sono due successioni crescenti in $\mathcal{A}$ tali che $\bigcup_n A_n \subset \bigcup_m A'_m$, allora $\lim_n P(A_n) \leq \lim_m P(A'_m)$. Vale l'uguale se $\bigcup_n A_n = \bigcup_m A'_m$.*

Dimostrazione. Fissato n abbiamo che $(A_n \cap A'_m)_m$ è crescente in $\mathcal{A}$ e $\bigcup_m (A_n \cap A'_m) = A_n$. Inoltre $\lim_m P(A'_m) \geq \lim_m P(A_n \cap A'_m) = P(A_n)$. Questo vale per ogni n. □

Vediamo ora le proprietà della classe $\mathcal{G}$ e della funzione Q appena definita.

Proposizione 1.1.4. *Siano Q e $\mathcal{G}$ come sopra. Valgono le seguenti proprietà:*

1. *$\emptyset, \Omega \in \mathcal{G}$ e $Q(\Omega) = 1, Q(\emptyset) = 0$. Inoltre $0 \leq Q(G) \leq 1$ per $G \in \mathcal{G}$;*
2. *Se $G_1, G_2 \in \mathcal{G}$ allora*
$$G_1 \cup G_2 \in \mathcal{G},$$
$$G_1 \cap G_2 \in \mathcal{G}$$
e
$$Q(G_1 \cup G_2) = Q(G_1) + Q(G_2) - Q(G_1 \cap G_2);$$
3. *Se $G_1, G_2 \in \mathcal{G}$ e $G_1 \subset G_2$, allora $Q(G_1) \leq Q(G_2)$*
4. *Se $G_n \to G$ e $G_n \in \mathcal{G}$, allora $G \in \mathcal{G}$ e $Q(G) = \lim_n Q(G_n)$.*

Inoltre $Q(A) = P(A)$ se $A \in \mathcal{A}$.

Dimostrazione.

1. è ovvia;
2. Consideriamo $(A'_n)_{n\geq 1}, (A_n)_{n\geq 1}$ successioni crescenti in $\mathcal{A}$ con limiti rispettivamente G', G. Sappiamo che $P(A'_n \cup A_n) = P(A'_n) + P(A_n) - P(A'_n \cap A_n)$ e facendo il limite otteniamo la tesi;
3. Segue dal lemma precedente;
4. Supponiamo che $(A_{m,n})_{m\geq 1}$ siano successioni crescenti in $\mathcal{A}$ con limiti rispettivamente G_n e che $G_n \nearrow G$. Sia $D_m = \sup_{n\leq m} A_{m,n}$; $D_m \in \mathcal{A}$ e $(D_m)_m$ è crescente. Abbiamo che $A_{m,n} \subset D_m \subset G_m$ e quindi $P(A_{m,n}) \leq P(D_m) \leq Q(G_m)$ per $n \leq m$. Mandando $m, n \to \infty$ otteniamo che $D_m \nearrow G \in \mathcal{G}$ e $Q(G) = \lim_m P(D_m) = \lim_m Q(G_m)$.

□

Proposizione 1.1.5. *Siano $\mathcal{G}$ e Q come prima. Allora Q_* definita sull'insieme delle parti di Ω come $Q_*(\Omega_1) = \inf_{\Omega_1\subset G\in\mathcal{G}} Q(G)$, ha le seguenti proprietà:*

1. *$Q_*(G) = Q(G)$ se $G \in \mathcal{G}$. Inoltre $0 \leq Q_*(\Omega_1) \leq 1$.*
2. *$Q_*(\Omega_1 \cup \Omega_2) \leq Q_*(\Omega_1) + Q_*(\Omega_2) - Q_*(\Omega_1 \cap \Omega_2)$. In particolare $Q_*(\Omega_1) + Q_*(\Omega_1^c) \geq 1$*
3. *Se $\Omega_1 \subset \Omega_2$ allora $Q_*(\Omega_1) \leq Q_*(\Omega_2)$*
4. *Se $\Omega_n \nearrow \Omega_\infty$ allora $Q_*(\Omega_n) \nearrow Q_*(\Omega_\infty)$*

Dimostrazione.

1. è banale
2. Fissiamo $\varepsilon > 0$ e siano $G_1, G_2 \in \mathcal{G}$ tali che $\Omega_i \subset G_i$ e $Q_*(\Omega_i) + \frac{\varepsilon}{2} \geq Q(G_i), i = 1, 2$. Allora $Q_*(\Omega_1) + Q_*(\Omega_2) + \epsilon \geq Q(G_1) + Q(G_2) = Q(G_1 \cup G_2) + Q(G_1 \cap G_2) \geq Q_*(\Omega_1 \cup \Omega_2) + Q_*(\Omega_1 \cap \Omega_2)$
3. Diretta conseguenza della monotonia di Q

4. Fissiamo $\varepsilon > 0$ e sia $\varepsilon_n > 0$ tale che $\sum_{n=1}^{\infty} \varepsilon_n = \varepsilon$ e $G_n \in \mathcal{G}$ tale che $\Omega_n \subset G_n$ e $Q_*(\Omega_n) + \varepsilon_n \geq Q(G_n)$. Sia $G'_n = \bigcup_{m \leq n} G_m$, così che $\Omega_n \subset G'_n$ e $(G'_n)_{n\geq 1}$ è una successione crescente in $\mathcal{G}$. Facciamo vedere per induzione che

$$Q_*(\Omega_n) + \sum_{m \leq n} \varepsilon_m \geq Q(G'_n).$$

Per $n = 1$ è vero per ipotesi. Supponiamo sia vero per n. Dimostriamo che vale per $n + 1$.

Poiché $\Omega_n \subset G'_n \cap G_{n+1} \in \mathcal{G}$, abbiamo

$$Q(G'_{n+1}) = Q(G'_n \cup G_{n+1}) = Q(G'_n) + Q(G_{n+1}) - Q(G'_n \cap G_{n+1}) \leq$$

$$\leq Q_*(\Omega_n) + \sum_{m \leq n} \varepsilon_m + Q_*(\Omega_{n+1}) + \varepsilon_{n+1} - Q_*(\Omega_n) =$$

$$= Q_*(\Omega_{n+1}) + \sum_{m \leq n+1} \varepsilon_m.$$

Allora vale anche per $n + 1$.

Mandiamo ora $n \to \infty$ e prendiamo $\Omega_\infty \subset \lim_n G'_n \in \mathcal{G}$. Abbiamo quindi $\lim Q_*(\Omega_n) + \varepsilon \geq Q(\lim G'_n) \geq Q_*(\Omega_\infty)$. Visto che ε è arbitrario e che per ogni $n > 0$

$$Q_*(\Omega_n) \leq Q_*(\Omega_\infty)$$

per la monotonia, allora abbiamo la tesi.

□

Corollario 1.1.6. *Nelle condizioni precedenti, la classe $\mathcal{D}$ di insiemi D in Ω tale che $Q_*(D) + Q_*(D^c) = 1$ è una σ-algebra. Inoltre Q_* ristretto a $\mathcal{D}$ è una probabilità completa su $(\Omega, \mathcal{D})$.*

Dimostrazione. Dalla definizione segue che se $D \in \mathcal{D}$, allora $D^c \in \mathcal{D}$. Inoltre dalla 1 viene che che $\emptyset \in \mathcal{D}$ e quindi $\Omega \in \mathcal{D}$. Se $D_1, D_2 \in \mathcal{D}$ abbiamo che $Q_*(D_1 \cup D_2) + Q_*(D_1 \cap D_2) \leq Q_*(D_1) + Q_*(D_2)$ e $Q_*((D_1 \cup D_2)^c) + Q_*((D_1 \cap D_2)^c) \leq Q_*(D_1^c) + Q_*(D_2^c)$ e poiché $Q_*(D_1 \cup D_2) + Q_*((D_1 \cup D_2)^c) \geq 1$ e

$Q_*(D_1 \cap D_2) + Q_*((D_1 \cap D_2)^c) \geq 1$ per forza vale l'uguale. Allora $D_1 \cup D_2 \in \mathcal{D}$ e $D_1 \cap D_2 \in \mathcal{D}$.

Se $(D_n)_{n\geq 1}$ è crescente in $\mathcal{D}$ abbiamo che $Q_*(\bigcup_n D_n) = \lim Q_*(D_n)$ e $Q_*((\bigcup_n D_n)^c) < Q_*(D_m^c)$ se $m \geq 1$. Questo implica che $Q_*(\bigcup_n D_n) + Q_*((\bigcup_n D_n)^c) \leq 1$ e quindi $\bigcup_n D_n \in \mathcal{D}$. Allora $\mathcal{D}$ è una σ-algebra e Q_* ristretta su di essa è una probabilità. Inoltre contiene tutti gli insiemi Q_*-nulli. □

Ora siamo pronti per dimostrare il teorema più importante della sezione.

Teorema 1.1.7. *Ogni probabilità P definita su un'algebra $\mathcal{A}$ di sottoinsiemi di Ω ha un'unica estensione sulla σ-algebra generata da $\mathcal{A}$*

Dimostrazione. Se P è una probabilità su $(\Omega, \mathcal{A})$ con $\mathcal{A}$ algebra, applichiamo il risultato del corollario precedente alla classe $\mathcal{G}$ e alla funzione Q definite come prima. Notiamo che Q e Q_* coincidono su $\mathcal{A}$ con P e da questo segue che $\mathcal{A} \subset \mathcal{D}$. Di conseguenza la restrizione P' di Q_* alla σ-algebra $\mathcal{A}'$ generata da $\mathcal{A}$, che è ancora contenuta in $\mathcal{D}$, è una estensione di P su questa σ-algebra.

Per mostrare l'unicità dell'estensione, consideriamo una probabilità Q' su $\mathcal{A}'$ la cui restrizione su $\mathcal{A}$ è uguale a P. $Q' = Q$ su $\mathcal{G}$ e quindi $Q' \leq P'$ su $\mathcal{A}'$. Se però $Q'(A) < P'(A)$ per ogni $A \in \mathcal{A}$ avremmo che $Q'(\Omega) = Q'(A) + Q'(A^c) < P'(A) + P'(A^c) = 1$ che è assurdo! Allora $Q'(A) = P'(A)$ □

1.2 Misura e integrazione

Definizione 1.4. (Misura)
Una misura positiva sulla σ-algebra $\mathcal{F}$ di Ω è un'applicazione P non negativa da $\mathcal{F}$ ai reali tale che

1. $P(\emptyset) = 0$

2. (sigma additività) per ogni successione $(A_n)_{n\in\mathbb{N}}$ di elementi di $\mathcal{F}$ a due a due disgiunti , $P(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} P(A_n)$

Se $P(\Omega) < +\infty$ la misura si dice finita. Se inoltre $P(\omega) = 1$ allora questa si dice misura di probabilità. Se P è una applicazione da $\mathcal{F}$ ai reali tale che valga solo la sigma additività allora si dice misura con segno.

Osservazione 4. Segue dalla definizione che se E è contenuto in G allora la misura di E è minore di quella di G.

Uno spazio di misura è la terna $(\Omega, \mathcal{F}, P)$. La coppia $(\Omega, \mathcal{F})$ viene chiamato spazio misurabile.

Definizione 1.5. (funzioni misurabili)
Dato uno spazio misurabile $(\Omega, \mathcal{F})$, f funzione da Ω in $\mathbb{R}^n$ è $\mathcal{F}$-misurabile se per ogni $H \in \mathcal{B}$ si ha $f^{-1}(H) \in \mathcal{F}$. In tal caso si scrive $f \in m\mathcal{F}$.

Nel caso particolare in cui $(\Omega, \mathcal{F}) = (\mathbb{R}^d, \mathcal{B})$, si dice che f è Borel-misurabile e si scrive $X \in m\mathcal{B}$.

Definiamo ora gli integrali data una misura generica

Definizione 1.6 (costruzione e definizione dell'integrale)**.** Il primo passo è definire l'integrale per una funzione h semplice, cioè $h = \sum_{k=1}^{n} \alpha_k I(A_k)$ con α_k reale e $I(A_k)$ la funzione caratteristica dell'insieme A_k. Definiamo

$$\int_\Omega h\, d\mu = \sum_{k=1}^{n} \alpha_k \mu(A_k).$$

Se questa somma vale $+\infty$ o $-\infty$ diremo che l'integrale non esiste.

Il secondo passo è definire l'integrale per una funzione h positiva borel-misurabile. Definiamo

$$\int_\Omega h\, d\mu = \sup\{\int_\Omega s\, d\mu : s \quad \textit{semplice}, \quad 0 \leq s \leq h\}.$$

Notiamo che questo integrale può valere anche ∞.

Il terzo passo si ha considerando una funzione h borel-misurabile non necessariamente positiva. In questo caso si considerano la parte positiva h^+ e la parte negativa h^-. Si definisce quindi

$$\int_\Omega h\, d\mu = \int_\Omega h^+\, d\mu - \int_\Omega h^-\, d\mu$$

a meno che non valga $+\infty - \infty$.

Se l'integrale è finito si dice che h è μ-integrabile o μ-sommabile.

L'integrale ha le proprietà di linearità e monotonia. Valgono inoltre i teoremi classici come Beppo-Levi, Fatou, Lebesgue.

1.3 Variabili aleatorie

Definizione 1.7. (Variabile aleatoria)
Dato uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, una variabile aleatoria(v.a.) X è una funzione da Ω in $\mathbb{R}^n$ tale che per ogni $H \in \mathcal{B}$ si ha $X^{-1}(H) \in \mathcal{F}$, cioè una v.a. è una funzione $\mathcal{F}$-misurabile.

Osservazione 5. Le v.a. si possono in realtà definire come $X : (\Omega, \mathcal{F}) \to (\tilde{\Omega}, \tilde{\mathcal{F}})$ misurabili, cioè per ogni $H \in \tilde{\mathcal{F}}$ si ha $X^{-1}(H) \in \mathcal{F}$.

Definizione 1.8. (σ-algebra generata da una v.a.)
Sia X una v.a. Definiamo la σ-algebra generata da X $\sigma(X) := X^{-1}(\mathcal{B}) = \{X^{-1}(H) | H \in \mathcal{B}\}$.

Questa è effettivamente una σ-algebra visto che presa una famiglia $(H_n)_{n\in\mathbf{I}}$ di sottoinsiemi di $\mathbb{R}^d$ ho che $\bigcup_{n\in\mathbf{I}} X^{-1}(H_n) = X^{-1}(\bigcup_{n\in\mathbf{I}} H_n)$ e preso $H \in \mathbb{R}^d$ ho $X^{-1}(H^c) = (X^{-1}(H))^c$.

Osservazione 6. $X \in m\mathcal{F}$ se e solo se $\sigma(X) \subset \mathcal{F}$.

Definizione 1.9. (Distribuzione)
Data una v.a. X si definisce $\pi_X : \mathcal{B} \to [0,1]$ con $\pi_X(H) = P^{-1}(H)$. π_X è chiamata distribuzione o legge di X e si scrive $X \sim \pi_X$.

Introduciamo ora uno dei concetti fondamentali associati a una v.a. X.

Definizione 1.10. (valore atteso)
Il valore atteso di una v.a.r. X è definito come $E[X] = \int_\Omega X\, dP$. Diciamo che X ha media se $\int_\Omega X\, dP$ non è indefinito.

Se X è un vettore aleatorio, cioè $X = (X_1, ..., X_d)$ a valori in R^d allora diremo che X ha media se ogni componente di X ha media e definiamo il suo valore attesto $E[X] = (E[X_1], ..., E[X_d])$.

Il valore atteso per come definito ha tutte le proprietà dell'integrale definito nella sezione precedente.

1.4 Probabilità condizionata e indipendenza

Definizione 1.11. (Probabilità condizionata)
Dati due eventi A e B con B non trascurabile. La probabilità di A condizionata a B è

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

Intuitivamente si può pensare che se accada B allora la probabilità che accada A possa cambiare.

Se B non influisce su A questi si diranno indipendenti, cioè A e B son indipendenti se $P(A|B) = P(A)$. Da questo si deduce che sono indipendenti se e solo se $P(A \cap B) = P(A)P(B)$.

Ora estendiamo a più eventi la definizione di indipendenza:

Definizione 1.12. (Indipendenza di eventi)
Sia $(A_i)_{i \in I}$ una famiglia finita o numerabile di eventi. Diciamo che tali eventi sono indipendenti se vale $P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$ per ogni $J \subset I$, con J finito.

Definizione 1.13. (Indipendenza di v.a.)
Siano X_i, $i = 1, ..., n$, variabili aleatorie definite sullo spazio di probabilità $(\Omega, \mathcal{F}, P)$, a valori rispettivamente in $\mathbb{R}^{d_i}$, $i = 1, ..., n$. Diciamo che X_i, $i = 1, ..., n$, sono variabili aleatorie indipendenti (nella misura di probabilità P) se, per ogni scelta dei sottoinsiemi $H_i \in B(\mathbb{R}^{d_i})$, $i = 1, ..., n$, gli eventi $X_1^{-1}(H_1), ..., X_n^{-1}(H_n)$ sono indipendenti. Più in generale, data una famiglia qualsiasi $(X_i)_{i \in I}$ di v.a. definite su $(\Omega, \mathcal{F}, P)$, esse si dicono indipendenti se ogni sottofamiglia finita $(X_i)_{i \in J}$, con $J \subset I$, è formata da v.a. indipendenti.

Capitolo 2

Attesa condizionata

2.1 Teorema di Radon-Nikodym

Date due misure Q e P su $(\Omega, \mathcal{F})$, diremo che Q è P-assolutamente continua su $\mathcal{F}$ se per ogni $A \in \mathcal{F}$ tale che $P(A) = 0$ si ha $Q(A) = 0$. In tal caso scriviamo $Q <<_{\mathcal{F}} P$. Viene messa in evidenza la σ-algebra perché l'assoluta continuità dipende da essa; se infatti $\mathcal{G} \subset \mathcal{F}$ sono due σ-algebre allora $Q <<_{\mathcal{G}} P$ non implica $Q <<_{\mathcal{F}} P$.

Se $Q <<_{\mathcal{F}} P$ e $P <<_{\mathcal{F}} Q$ diciamo che le misure P e Q sono equivalenti e scriviamo $P \sim Q$.

Osservazione 7. Se $Q <<_{\mathcal{F}} P$ allora gli eventi di $\mathcal{F}$ trascurabili per P lo sono anche per Q, ma non è detto il viceversa. Inoltre se $Q <<_{\mathcal{F}} P$ allora gli eventi certi per P lo sono anche per Q, ma come prima il viceversa può non valere.

Teorema 2.1.1. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio con misura finita. Se Qè una misura finita su $(\Omega, \mathcal{F})$ e $Q <<_{\mathcal{F}} P$, allora esiste $f : \Omega \to [0, +\infty[$ tale che*

1. *f è $\mathcal{F}$-misurabile;*

2. *f è P-sommabile;*

3.

$$Q(A) = \int_A f \, dP \quad \forall A \in \mathcal{F}.$$

Inoltre f è unica P-q.c. Si dice che f è la densità di Q rispetto a P su $\mathcal{F}$.

Per dimostrare questo teorema mostriamo prima alcuni risultati.

Lemma 2.1.2. *Sia Q una misura con segno. Se $E_n \nearrow E$ o $E_n \searrow E$, allora $Q(E_n) \to Q(E)$.*

Dimostrazione. Se $E_n \nearrow E$, allora $Q(E) = Q(E_1 \cup \bigcup_{n=1}^{\infty}(E_{n+1} \setminus E_n)) = Q(E_1) + \sum_{n=1}^{\infty}(E_{n+1} \setminus E_n) = \lim_i [Q(E_1) + \sum_{n=1}^{i-1} Q(E_{n+1} \setminus E_n)] = \lim_i Q(E_i)$. Se $E_n \searrow E$, allora $E_n^c \nearrow E^c$, e quindi $Q(E_n) = Q(\Omega) - Q(E_i^c) \to Q(\Omega) - Q(E^c) = Q(E)$. □

Teorema 2.1.3 (Decomposizione di Hahn). *Dato uno spazio misurabile $(\Omega, \mathcal{F})$ e una misura con segno Q, esistono due insiemi A^+ ed A^- in Ω tale che*

1. *$A^+ \cup A^- = \Omega$ e $A^+ \cap A^- = \emptyset$;*
2. *per ogni $E \in \Omega$ tale che $E \subset A^+$ abbiamo $Q(E) \geq 0$(A^+ si dice positivo);*
3. *per ogni $E \in \Omega$ tale che $E \subset A^-$ abbiamo $Q(E) \leq 0$(A^- si dice negativo).*

Dimostrazione. Prendiamo $\alpha = sup_{A \subset \mathcal{F}} Q(A)$.

Supponiamo esista un insieme A^+ tale che $Q(A^+) = \alpha$ e consideriamo $A^- = \Omega \setminus A^+$.

Se $A \subset A^+$ e $Q(A) < 0$ si ha $Q(A^+ \setminus A) > \alpha$ che è impossibile. Questo implica allora che A^+ è un insieme positivo.

Se $A \subset A^-$ e $Q(A) > 0$ si ha $Q(A^+ \cup A) > \alpha$ che è impossibile. Abbiamo quindi che A^- è un insieme negativo.

Dobbiamo quindi riusciamo a costruire un insieme A^+ tale che $Q(A^+) = \alpha$. Scegliamo (A_n) tale che $Q(A_n) \to \alpha$ e prendiamo $A = \bigcup_n A_n$. Per ogni n consideriamo 2^n insiemi $B_{n,i}$ della forma $\bigcap_{k=1}^n A'_k$ dove A'_k sono o A_k oppure $A \setminus A_k$.

La famiglia $\mathcal{B}_n = \{B_{n,i} : 1 \leq i \leq 2^n\}$ è una partizione di A. Notiamo che ogni $B_{n,j}$ è contenuto esattamente in uno dei $B_{n-1,i}$.

Sia $C_n = \cup_i B_{n,i}$ in $\mathcal{B}_n$ per cui $Q(B_{n,i}) > 0$. Visto che A_n è l'unione di alcuni $B_{n,i}$ abbiamo che $Q(A_n) \leq Q(C_n)$. Poiché i $\mathcal{B}_n$ sono sempre più fini, $m < n$ implica che $(\cup_{i=m}^n C_i) \setminus (\cup_{i=m}^{n-1} C_i)$ è l'unione(forse vuota) di alcuni insiemi $B_{n,i}$; i $B_{n,i}$ in questa unione son tale che $Q(B_{n,i}) > 0$ poiché contenuti in C_n.

Allora $Q(\cup_{i=m}^{n-1} C_i) \leq Q(\cup_{i=m}^n C_i)$, così per induzione ho che $Q(A_m) \leq Q(C_m) \leq Q(\cup_{i=m}^n C_i)$. Preso $D_m = \cup_{n=m}^\infty C_n$, dal lemma precedente ho $Q(A_m) \leq Q(D_m)$. Sia $A^+ = \cap_m^\infty D_m$, così che $D_m \searrow A^+$. Dal lemma precedente abbiamo che $\alpha = \lim_m Q(A_m) \leq \lim_m Q(D_m) = Q(A^+)$. Allora Q assume il suo massimo valore in A^+. □

Ora abbiamo tutti gli elementi per poter dimostrare il teorema di Radon-Nikodym.

Dimostrazione. Siano P e Q misure finite non negative, e sia F l'insieme delle funzioni misurabili $f : X \to [0, +\infty[$ che soddisfano $\int_A f\, dP \leq Q(A) \forall A \in \mathcal{F}$.

L'insieme F non è vuoto, poiché contiene almeno la funzione nulla. Siano $f_1, f_2 \in F$, A un insieme misurabile e

$A_1 = \{x \in A : f_1(x) > f_2(x)\} \qquad A_2 = \{x \in A : f_2(x) \geq f_1(x)\}$.

Allora abbiamo

$$\int_A \max(f_1, f_2)\, dP = \int_{A_1} f_1\, dP + \int_{A_2} f_2\, dP \leq Q(A_1) + Q(A_2) = Q(A)$$

e dunque $\max(f_1, f_2) \in F$

Sia ora (f_n) una successione di funzioni in F tali che

$$\lim_{n\to\infty} \int_\Omega f_n\, dP = \sup_{f \in F} \int_\Omega f\, dP.$$

Sostituendo f_n con il max delle prime n funzioni possiamo assumere che la successione (f_n) sia crescente. Sia g la funzione definita come

$$g(x) := \lim_{n\to\infty} f_n(x).$$

Per mostrare che g è la funzione cercata, cioè che il suo integrale su A rispetto a P vale esattamente $Q(A)$, notiamo che dal teorema della convergenza monotona

$$\int_A g\,dP = \lim_{n\to\infty}\int_A f_n\,dP \le Q(A) \qquad \forall A \in \mathcal{F}$$

e quindi $g \in F$. Inoltre, dalla costruzione di g segue

$$\int_\Omega g\,dP = \sup_{f\in F}\int_\Omega f\,dP.$$

Poiché $g \in F$, $Q_0(A) := Q(A) - \int_A g\,dP$ definisce una misura non negativa su $\mathcal{F}$. Supponendo quindi per assurdo $Q_0 \neq 0$, dato che P è finita esiste un $\varepsilon > 0$ tale che $Q_0(\Omega) > \varepsilon P(\Omega)$.

Sia allora (A^+, A^-) la decomposizione di Hahn per la misura con segno $Q_0 - \varepsilon P$. Per ogni $A \in \mathcal{F}$ abbiamo $Q_0(A \cap A^+) \ge \varepsilon P(A \cap A^+)$ e quindi

$$Q(A) = \int_A g\,dP + Q_0(A) \ge \int_A g\,dP + Q_0(A\cap A^+) \ge$$

$$\ge \int_A g\,dP + \varepsilon P(A\cap A^+) = \int_A (g + \varepsilon \mathbb{1}_{A^+})\,dP$$

Notiamo che $P(A^+) > 0$; se fosse nulla, poiché Q è assolutamente continua rispetto a P avremmo $Q_0(A^+) \le Q(A^+) = 0$, quindi $Q_0(A^+) = 0$ e $Q_0(\Omega) - \varepsilon P(\Omega) = (Q_0 - \varepsilon P)(A^-) \le 0$ contraddicendo il fatto che $Q_0(\Omega) > \varepsilon P(\Omega)$.

Essendo che $\int_\Omega (g + \varepsilon \mathbb{1}_{A^+})\,dP \le Q(\Omega) < +\infty$ la funzione $g + \varepsilon \mathbb{1}_{A^+} \in F$ e soddisfa

$$\int_\Omega (g + \varepsilon \mathbb{1}_{A^+})\,dP > \int_\Omega g\,dP = \sup_{f\in F}\int_\Omega f\,dP$$

ma questo è impossibile, e quindi l'assunzione iniziale che $Q_0 \neq 0$ deve essere falsa.

Dato che g è P-integrabile, l'insieme$\{x \in \Omega : g(x) = +\infty\}$ è P-nullo. Quindi f è definita come $f(x) = \begin{cases} g(x) & \text{se } g(x) < \infty \\ 0 & \text{altrimenti} \end{cases}$ e possiede le proprietà richieste.

Per l'unicità siano f, g due funzioni misurabili che soddisfano $Q(A) = \int_A f\, dP = \int_A g\, dP$ per ogni A misurabile. $f - g$ è P-integrabile e $\int_A (f - g)\, dP = 0$, allora $f = g$ P-q.c. $\square$

2.2 Esistenza e proprietà

Definizione 2.1. (attesa condizionata)
Sia X una v.a.r, integrabile su $(\Omega, \mathcal{F}, P)$ e $\mathcal{G} \subset \mathcal{F}$ una σ-algebra. Chiamiamo attesa condizionata o speranza condizionata di X rispetto a $\mathcal{G}$ ogni variabile aleatoria X' $\mathcal{G}$-misurabile tale che

$$\int_B X\, dP = \int_B X'\, dP \quad \forall B \in \mathcal{G}.$$

Chiameremo con lo stesso nome la classe di P-equivalenza di tali v.a. L'attesa condizionata viene denotata con $E[X|\mathcal{G}]$. Ogni uguaglianza fra attesa condizionata e v.a. sarà sempre intesa P-q.c. oppure come classi di P-equivalenza.

Come nel caso della probabilità condizionata, anche qui intuitivamente si può pensare che avendo a disposizione le informazioni della sigma algebra $\mathcal{G}$ la nostra attesa condizionata sia più corretta della semplice $E[X]$.

Lo scopo dell'osservazione seguente è quello di motivare tale definizione.

Osservazione 8. Sia X una v.a.r. sommabile e B un evento con probabilità $P(B) > 0$. L'attesa di X condizionata all'evento B è definita come l'attesa condizionata di X rispetto alla misura $P(\cdot|B)$, cioè

$$E[X|B] = \frac{1}{P(B)} \int_B X\, dP.$$

Sia $\mathcal{G}$ la σ-algebra generata da B, cioè $\mathcal{G} = \{\emptyset, \Omega, B, B^c\}$.

L'attesa di X condizionata a $\mathcal{G}$ è è definita da

$$E[X|\mathcal{G}](\omega) = \begin{cases} E[X|B] & \text{se } \omega \in B \\ E[X|B^c] & \text{se } \omega \in B^c \end{cases}$$

$E[X|\mathcal{G}]$ è una variabile aleatoria e si vede che verifica le proprietà richieste dalla definizione.

In modo più generale, prendendo $\mathcal{G}$ generata da una partizione finita di $\{B_1, ...B_n\}$ abbiamo che

$$E[X|\mathcal{G}] = \sum_{i=1}^{n} (\frac{1}{P(B_i)} \int_{B_i} X\, dP) \mathbb{1}_{B_i}.$$

Dimostrazione. $X' = E[X|\mathcal{G}] = \sum_{i=1}^{n} (\frac{1}{P(B_i)} \int_{B_i} X\, dP) \mathbb{1}_{B_i}$ è una v.a. $\mathcal{G}$-misurabile, infatti presa la controimmagine di un boreliano H questa può appartenere o a $\emptyset$,nel caso in cui nessun valore di X' appartenga ad H, oppure all'unione di qualche B_i.

Inoltre, visto che $G \in \mathcal{G}$ può essere scritto come unione di B_i,

$$E[X \mathbb{1}_G] = \sum_{B_i \in G} \int_{B_i} X\, dP$$

e

$$E[X' \mathbb{1}_G] = \sum_{i=1}^{n} (\frac{1}{P(B_i)} \int_{B_i} X\, dP) E[\mathbb{1}_{B_i} \mathbb{1}_G] =$$

$$\sum_{i=1}^{n} E[X|B_i] E[\mathbb{1}_{B_i \cap G}] = \sum_{i=1}^{n} E[X|B_i] P(B_i \cap G) = \sum_{B_i \in G} \int_{B_i} X\, dP$$

$\square$

Osservazione 9. La definizione data è equivalente a chiedere che X' sia $\mathcal{G}$-misurabile e che valga $E[XY] = E[X'Y]$ per ogni Y v.a. limitata $\mathcal{G}$-misurabile.

Dimostrazione. Prendendo Y della forma $\mathbb{1}_B$ con $B \in \mathcal{G}$ si ha la prima implicazione. Viceversa, possiamo riscrivere l'uguaglianza integrale come $E[X \mathbb{1}_B] = E[X' \mathbb{1}_B]$. Discende poi dalla definizione che questa varrà per ogni Y combinazione lineare di funzioni indicatrici, cioè $E[XY] = E[X'Y]$. Ora basta approssimare puntualmente dal basso ogni v.a. limitata $\mathcal{G}$-misurabile e passare al limite. $\square$

Viene spontaneo porsi la domanda se possa esistere una v.a. con tali proprietà.

Teorema 2.2.1 (Esistenza dell'attesa condizionata)**.** *Data una v.a. a valori reali integrabile su* $(\Omega, \mathcal{F}, P)$ *ed una* σ*-algebra* $\mathcal{G} \subset \mathcal{F}$*, esiste sempre una v.a.* X' $\mathcal{G}$*-misurabile e tale che* $\int_B X\, dP = \int_B X'\, dP$ *per ogni* $B \in \mathcal{G}$*. Inoltre questa è unica a meno di* P*-equivalenza.*

Dimostrazione. Dimostriamo inizialmente il teorema per una v.a. X non negativa. X è sommabile e per questo $Q(B) = \int_B X\, dP$ definisce una misura finita Q su $\mathcal{G}$. Inoltre $Q <<_{\mathcal{G}} P$ e quindi per il teorema di Radon-Nikodym esiste Y v.a. $\mathcal{G}$-misurabile tale che $Q(B) = \int_B Y\, dP$ per ogni $B \in \mathcal{G}$. Per il caso generale scomponiamo $X = X^+ - X^-$. $E[X|F] = E[X^+|F] - E[X^-|F]$ soddisfa entrambe le condizioni della definizione. Con questo l'esistenza è dimostrata.

Unicità: Prendiamo Y_1 e Y_2 v.a. che soddisfano le condizioni della definizione. $B = \{Y_1 > Y_2\}$ è un evento di $\mathcal{G}$. Ho per definizione $\int_B (Y_1 - Y_2)\, dP = \int_B X\, dP - \int_B X\, dP = 0$. Questo implica che $P(B) = 0$ e per simmetria ho che $Y_1 = Y_2$ q.c. $\square$

Proposizione 2.2.2 (Proprietà)**.** *Siano* X*,* Y *integrabili,* $\mathcal{G}$ *una sotto* σ*-algebra di* $\mathcal{F}$*. Valgono le seguenti affermazioni:*

1. *Se* X *è* $\mathcal{G}$*-misurabile allora* $X = E[X|\mathcal{G}]$;
2. *Se* $\mathcal{G}' \subset \mathcal{G} \subset \mathcal{F}$ *allora* $E[E[X|\mathcal{G}]|\mathcal{G}'] = E[X|\mathcal{G}']$*;*
3. *Se* Y *è* $\mathcal{G}$*-misurabile ed* YX *è integrabile, allora* $E[YX|\mathcal{G}] = YE[X|\mathcal{G}]$*;*
4. *Se* X *è indipendente da* $\mathcal{G}$ *allora* $E[X|\mathcal{G}] = E[X]$*;*
5. $E[aX + bY + c|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}] + c$*;*
6. *Se* $X \leq Y$ *q.c. allora* $E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$*;*
7. *Se* Y *è indipendente da* $\sigma(X, \mathcal{G})$ *allora* $E[XY|\mathcal{G}] = E[X|\mathcal{G}]E[Y]$*;*
8. *Se* $(X_n)_{n\in\mathbb{N}}$ *è una successione di v.a. non negative integrabili e monotona crescente, con* $X = \lim_{n\to\infty} X_n$ *integrabile, allora* $E[X_n|\mathcal{G}] \to E[X|\mathcal{G}]$ *q.c.*

9. *Se* $(X_n)_{n\in\mathbb{N}}$ *è una successione di v.a. non negative sommabili allora posto* $X = \liminf X_n$ *ho* $\liminf E[X_n|\mathcal{G}] \geq E[X|\mathcal{G}]$*;*

10. *Se* $(X_n)_{n\in\mathbb{N}}$ *una successione di v.a. convergente puntualmente a* X *q.c. e se esiste* Y *sommabile tale che* $|X_n| \leq Y$ *q.c. allora* $\lim_{n\to\infty} E[X_n|\mathcal{G}] = E[X|\mathcal{G}]$*;*

11. *(Jensen) Se* ϕ *è una funzione convessa tale che* $\phi(X)$ *sia sommabile, allora* $E[\phi(X)|\mathcal{G}] \geq \phi(E[X|\mathcal{G}])$*.*

Dimostrazione. Per dimostrare queste proprietà dobbiamo verificare le due proprietà della definizione oppure la definizione equivalente dell'osservazione.

1. X è $\mathcal{G}$-misurabile e quindi la prima proprietà è verificata. La seconda è ovvia.

2. $E[E[X|\mathcal{G}]|\mathcal{G}']$ è $\mathcal{G}'$-misurabile; inoltre se W è $\mathcal{G}'$-misurabile limitata è anche $\mathcal{G}$-misurabile limitata e vale $E[WE[X|\mathcal{G}]|\mathcal{G}'] = E[WE[X|\mathcal{G}]] = E[WX]$.

3. $YE[X|\mathcal{G}]$ è $\mathcal{G}$-misurabile e per ogni W limitata $\mathcal{G}$-misurabile vale $E[(WY)E[X|\mathcal{G}]] = E[WYX]$ essendo WY $\mathcal{G}$-misurabile e limitata.

4. $E[X]$ è $\mathcal{G}$-misurabile e vale l'uguaglianza integrale.

5. Le costanti sono $\mathcal{G}$-misurabili. Inoltre anche somma e prodotti di variabili aleatorie $\mathcal{G}$-misurabili lo sono, quindi abbiamo la prima proprietà. L'uguaglianza integrale segue immediatamente dalla linearità dell'integrale e per il fatto che $\int_B X\,dP = \int_B E[X|\mathcal{G}]\,dP$ e $\int_B Y\,dP = \int_B E[Y|\mathcal{G}]\,dP$.

6. Segue immediatamente dalla definizione.

7. $E[X|\mathcal{G}]E[Y]$ è $\mathcal{G}$-misurabile e per ogni W $\mathcal{G}$-misurabile limitata vale $E[WE[X|\mathcal{G}]E[Y]] = E[WE[X|\mathcal{G}]]E[Y] = E[WX]E[Y] = E[WXY]$.

8. Per la monotonia ho che $E[X_n|\mathcal{G}]$è una successione crescente di v.a. $\mathcal{G}$-misurabili e non-negative, quindi anche il sup è una v.a. $\mathcal{G}$-misurabile. Inoltre per ogni $B \in \mathcal{G}$ applicando Beppo-Levi abbiamo $\int_B sup\, dP = \lim_{n\to\infty} \int_B E[X_n|\mathcal{G}]\, dP = \lim_{n\to\infty} \int_B X_n\, dP = \int_B X\, dP$.

9. Sia $Y_k = \inf_{n\geq k} X_k \forall k \in \mathbb{N}$. Abbiamo ora una successione non negativa, crescente il cui limite è X. Inoltre per la monotonia $E[Y_k|\mathcal{G}] \leq E[X_n|\mathcal{G}]$ q.c. da cui $E[Y_k|\mathcal{G}] \leq \inf_{n\geq k} E[X_n|\mathcal{G}]$ q.c. Allora per il punto precedente e la disuguaglianza appena mostrata $E[\liminf X_n|\mathcal{G}] = E[X|\mathcal{G}] = E[\lim Y_k|\mathcal{G}] = \lim E[Y_k|\mathcal{G}] \leq \liminf E[X_n|\mathcal{G}]$

10. Applicando la proprietà precedente alla successione $Y - X_n$ ottengo $E[\limsup X_n|\mathcal{G}] \geq \limsup E[X_n|\mathcal{G}]$, da cui segue la proprietà desiderata;

11. Ogni funzione convessa coincide con l'inviluppo superiore delle funzioni lineari $l \leq \phi$, ossia
$$\phi(x) = \sup_{l\in L} l(x)$$
con $L = \{l : R \to R|l(x)\ lineare\ e\ l \leq \phi\}$. Si ha quindi $E[\phi(X)|\mathcal{G}] = E[\sup_{l\in L} l(X)|\mathcal{G}] \geq \sup_{l\in L} E[l(X)|\mathcal{G}] = \sup_{l\in L} l(E[X|\mathcal{G}]) = \phi(E[X|\mathcal{G}])$.

□

Osservazione 10. Se $\mathcal{N} = \{\Omega, \emptyset\}$ è la σ-algebra banale si ha $E[X|\mathcal{N}] = E[X]$. Possiamo quindi vedere l'attesa condizionata come una generalizzazione di quella definita nel capitolo 1.

Notazione. Se Y è una v.a. si scrive $E[X|Y]$ al posto di $E[X|\sigma(Y)]$

L'attesa condizionata può essere vista anche come una proiezione negli spazi L^p. Vale infatti la seguente proposizione

Proposizione 2.2.3. *L'attesa condizionata $E[\cdot|\mathcal{G}]$ ristretta allo spazio $L^p(\Omega, \mathcal{F}, P)$ è per ogni $p \geq 1$ una proiezione dello spazio $L^p(\Omega, \mathcal{F}, P)$ sullo spazio $L^p(\Omega, \mathcal{G}, P)$. In particolare sullo spazio $L^2(\Omega, \mathcal{F}, P)$, essendo uno spazio di Hilbert, è una proiezione ortogonale.*

Dimostrazione. Per la disuguaglianza di Jensen con la funzione convessa $\phi(x) = |x|^p$ abbiamo che $E[|X|^p|\mathcal{G}] \leq E[|X||\mathcal{G}]^p$ e $E[|X||\mathcal{G}] \leq |E[X|\mathcal{G}]|$. Questo implica che $E[|X|]^p = E[E[|X|^p|\mathcal{G}]] \geq E[|E[X|\mathcal{G}]|^p]$, cioè $||X||_p \geq ||E[X|\mathcal{G}]||_p$ per ogni $X \in L^p(\Omega, \mathcal{F}, P)$. Inoltre per la linearità e la idempotenza dimostrate prima abbiamo la tesi. □

Il precedente risultato ci permette di dare una nuova interpretazione all'attesa condizionata: può essere vista come la migliore approssimazione di X in $L^p(\Omega, \mathcal{G}, P)$.

Capitolo 3

Versione regolare della probabilità condizionata

3.1 Probabilità condizionata a una σ-algebra

Definizione 3.1. (Probabilità condizionata a una σ-algebra)
Sia $\mathcal{G}$ una sotto σ-algebra di $\mathcal{F}$ e sia $A \in \mathcal{F}$; si dice probabilità condizionata di A rispetto a $\mathcal{G}$ la classe di equivalenza di v.a.r. $\mathcal{G}$-misurabili

$$P(A|\mathcal{G}) = E[\mathbb{1}_A|\mathcal{G}].$$

Questa definizione si può motivare notando che nel caso particolare in cui $\mathcal{G}$ sia generata da $(B_1, ..., B_n)$ eventi, vale

$$E[\mathbb{1}_A|\mathcal{G}] = \sum_{i=1}^{n} P(A|B_i)\mathbb{1}_{B_i}$$

infatti

$$\int_{B_i} \mathbb{1}_A \, dP = P(A \cap B_i).$$

Ma questa sommatoria ha come termini le probabilità condizionate ai singoli eventi della σ-algebra; è spontaneo definire quindi

$$P(A|\mathcal{G}) = \sum_{i=1}^{n} P(A|B_i)\mathbb{1}_{B_i}$$

ottenendo l'uguaglianza della definizione. Inoltre in questo caso vale

$$P(A \cap B) = \int_B P(A|\mathcal{G})\, dP \quad \forall B \in \mathcal{G}.$$

Grazie alla definizione data viene generalizzata questa uguaglianza, infatti si ha

$$\int_B P(A|\mathcal{G})\, dP = \int_B \mathbb{1}_A\, dP = P(A \cap B).$$

Osservazione 11. Notiamo che $P(A|\mathcal{G}) \geq 0$ poiché la v.a. indicatrice di A è sempre positiva. Inoltre dalla proprietà di monotonia dell'attesa condizionata a una σ-algebra segue che $P(A|\mathcal{G}) \leq P(\Omega|\mathcal{G})$ e poiché $\mathbb{1}_\Omega$ è una v.a. costante uguale a 1 ho che $P(A|\mathcal{G}) \leq 1$ e $P(\Omega|\mathcal{G}) = 1$. Tutte queste proprietà sono verificate a meno di insiemi di probabilità P nulla.

Proposizione 3.1.1 (additività). *Sia $(A_n)_{n \in I}$ una famiglia finita di eventi a due a due disgiunti. Allora $P(\bigcup_n A_n|\mathcal{G}) = \sum_n P(A_n|\mathcal{G})$ q.c.*

Dimostrazione. Visto che gli eventi sono a due a due disgiunti ho che $\mathbb{1}_{\cup_n A_n} = \sum_n \mathbb{1}_{A_n}$ e grazie alla linearità dell'attesa condizionata ho la tesi. □

Si nota che la probabilità condizionata a una σ-algebra è definita a meno di eventi trascurabili. Non ha quindi senso fissare $\omega \in \Omega$ e considerare la funzione $A \to P(A|\mathcal{G})(\omega)$. Inoltre anche le proprietà precedenti sono verificate a meno di insiemi di misura nulla, che, in generale, dipendono dalla famiglia di eventi $(A_n)_{n \in \mathbb{N}}$ o dall'evento A. La domanda che sorge spontanea è dunque se è possibile far sì che fissato ω la funzione $A \to P(A|\mathcal{G})(\omega)$ sia una misura di probabilità.

3.2 Versione regolare della probabilità condizionata

Definizione 3.2. (versione regolare della probabilità condizionata)
Dati $(\Omega, \mathcal{F}, P)$ e $\mathcal{G} \subset \mathcal{F}$, una funzione $Q(\omega, A)$ definita su $\Omega \times \mathcal{F}$ si dice versione regolare della probabilità condizionata su $\mathcal{F}$ rispetto a $\mathcal{G}$ se

1. per ogni $A \in \mathcal{F}$ la funzione $\omega \to Q(\omega, A)$ sia misurabile ed appartenga alla classe di equivalenza di $P(A|\mathcal{G})$, cioè se $Q(\cdot, A)$ è una versione di $P(A|\mathcal{G})$;

2. per ogni $\omega \in \Omega$ $A \to Q(\omega, A)$ sia una misura di probabilità su $(\Omega, \mathcal{F})$.

Esempio 3.1. *Sia $\mathcal{G}$ la σ-algebra generata da B, cioè $\mathcal{G} = \{\emptyset, \Omega, B, B^c\}$. Avevamo già studiato l'attesa condizionata nel capitolo precedente. Prendiamo ora la funzione*

$$Q(\omega, G) = \mathbb{1}_B P(G|B) + \mathbb{1}_{B^c} P(G|B^c).$$

Verifichiamo che questa sia una versione regolare della probabilità condizionata. Consideriamo per ogni $G \in \mathcal{G}$ la funzione $\omega \to Q(\omega, G)$. Se $\omega \in B$ allora $\int_B Q(\omega, G)\, dP = \int_B P(G|B)\, dP = P(G|B)P(B) = P(G \cap B)$, ma $P(G|\mathcal{G})$ su B vale proprio $\frac{P(G\cap B)}{P(B)}$ e quindi $\int_B P(G|\mathcal{G})\, dP = \int_B Q(\omega, G)\, dP$. Lo stesso vale per gli altri insiemi contenuti in $\mathcal{G}$. Questo dimostra che $\omega \to Q(\omega, G)$ è una versione di $P(G|\mathcal{G})$. Ora facciamo vedere che per ogni $\omega \in \Omega$ $A \to Q(\omega, A)$ è una misura di probabilità. Le tre proprietà derivano subito dal fatto che fissato ω $Q(\omega, G)$ diventa una probabilità condizionata ad un evento.

Il primo problema che ci si pone è dove può esistere una versione regolare della probabilità. Affronteremo quindi il problema della sua esistenza. Consideriamo il solito spazio di probabilità $(\Omega, \mathcal{F}, P)$ e consideriamo $\mathcal{G}$ una sotto-σ-algebre di $\mathcal{F}$.

Definizione 3.3. (Classe compatta)
Una famiglia di insiemi $\mathcal{K} \subset \mathcal{F}$ si dice classe compatta se gode della proprietà dell'intersezione finita, cioè se per ogni $(K_n)_{n\in\mathbb{N}}$ di elementi di $\mathcal{K}$ tale che $\bigcap_{n=1}^{\infty} K_n = \emptyset$ esiste $m \in \mathbb{N}$ tale che$\bigcap_{n=1}^{m} K_n = \emptyset$.

Lemma 3.2.1. *Se $\mathcal{K}$ è una classe compatta, allora lo è anche la classe $\mathcal{K}'$ generata da $\mathcal{K}$, cioè chiusa rispetto all'unione finita e intersezione numerabile.*

Dimostrazione. Per prima cosa proviamo che la classe K_s delle unioni di famiglie finite di insiemi di $\mathcal{K}$ è compatta. Consideriamo $D_n = \bigcup_{m=1}^{M_n} K_n^m$ tale che $\bigcap_{n<p} D_n \neq \emptyset$ per ogni $p > 0$ e mostriamo che $\bigcap_n D_n \neq \emptyset$.

Formiamo i sottoinsiemi J_p, nello spazio $J = \prod_n \{1, ..., M_n\}$ di tutte le successioni $(m_n)_{n\geq 1}$ di interi tale che $1 \leq m_n \leq M_n$, di successioni $(m_n)_{n\geq 1}$ tale che $\bigcap_{n\leq p} K_n^{m_n} \neq \emptyset$.

Sappiamo che $\bigcap_{n\leq p} D_n = \bigcup_J (\bigcap_{n\leq p} K_n^{m_n})$, allora $J_p \neq \emptyset$ per ogni $p > 0$. I J_p sono inoltre decrescenti, quindi è sufficiente provare l'esistenza di una successione (m_n^*) nel $\lim_{p\to\infty} J_p$ per mostrare che $\bigcap_n K_n^{m_n^*} \subset \bigcap_n D_n \neq \emptyset$, poiché $\bigcap_{n\leq p} K_n^{m_n^*} \neq \emptyset$ e K è compatto.

Ma, dopo aver scelto una successione (m_n^q) in ogni insieme J_p, è possibile determinare per induzione su p, una successione (m_p^*) tale che per ogni p ci sono infiniti q per cui $m_n^q = m_n^*(1 \leq n \leq p)$.

Per p fissato, scegliamo q grande almeno come p. Allora la successione (m_n^q) appartiene a J_p, come (m_n^*) poiché nella definizione di J_p entrano solo i primi p termini. Abbiamo quindi mostrato che (m_n^*) appartiene a J_p per ogni p. Così K_s è compatta. Per finire è immediato che anche la classe delle intersezioni di famiglie numerabili di K_s è compatta. Inoltre questa coincide con K'. □

Definizione 3.4. Una σ-algebra si dice numerabilmente generata se e solo se esiste una sottoclasse $G' \subset \mathcal{G}$ numerabile e tale che la più piccola σ-algebra contenente G' coincida con $\mathcal{G}$.

Definizione 3.5. Una misura positiva P su $\mathcal{G}$ si dice regolare se esiste una classe compatta $\mathcal{K} \subset \mathcal{G}$ tale che per ogni $A \in \mathcal{G}$ si abbia

$$P(A) = \sup_{K\in\mathcal{K}, K\subset A} P(K). \tag{3.1}$$

Vale il seguente teorema di esistenza

Teorema 3.2.2. *Se $\mathcal{F}$ è numerabilmente generata e P è regolare su $\mathcal{F}$, allora esiste una versione regolare della probabilità condizionata su $\mathcal{F}$ rispetto a una qualunque sotto-σ-algebra $\mathcal{G}$.*

Dimostrazione. Sia $\{B_n\}_{n\geq 1}$ una famiglia numerabile che genera $\mathcal{F}$. Possiamo supporre che la classe $\mathcal{K}$ sia chiusa rispetto l'unione finita per il lemma precedente, quindi per ogni $n \geq 1$ esiste una successione crescente di sottoinsiemi di B_n in $\mathcal{K}$, che chiamiamo $\{C_n^k\}_{k\geq 1}$, tale che $P(B_n) = \lim_{k\to\infty} P(C_n^k)$. L'algebra $\mathcal{D}$ generata dai B_n e dai C_n^k è una sottoalgebra numerabile di $\mathcal{F}$.

Consideriamo ora una versione della probabilità condizionata alla σ-algebra $\mathcal{G}$ che denotiamo $Q(\cdot, A)$ con $A \in \mathcal{F}$. Sappiamo che valgono le seguenti proprietà:

1. per ogni $D \in \mathcal{D}$ abbiamo $0 \leq Q(\cdot, D) \leq 1$ al di fuori di un insieme P-trascurabile.

2. $Q(\cdot, \Omega) = 1$ al di fuori di un insieme P-trascurabile.

3. per ogni coppia D, D' di insiemi disgiunti in $\mathcal{D}$ $Q(\cdot, D \cup D') = Q(\cdot, D) + Q(\cdot, D')$ al di fuori di un insieme P-trascurabile.

4. per ogni $n \geq 1$ esiste un insieme P-trascurabile al di fuori del quale la successione crescente $Q(\cdot, C_n^k) \to Q(\cdot, B_n)$ per $k \to \infty$ (infatti $\mathbb{1}_{C_n^k} \nearrow \mathbb{1}_{B_n} P$-q.c.).

Queste proprietà valgono a meno di una famiglia numerabile di insiemi P-trascurabili, esiste quindi un insieme $\mathcal{N}$ P-trascurabile al di fuori del quale valgono le proprietà precedenti.

Ora guardiamo i due casi:

Se $\omega \notin \mathcal{N}$, allora restringendo $Q(\omega, \cdot)$ all'algebra $\mathcal{D}$ facciamo vedere che è una probabilità. Dobbiamo dimostrare che presa una successione $(D_n)_{n\in\mathbb{N}}$ decrescente a $\emptyset$ allora $Q(\omega, D_n) \to 0$. Siano $(C_n)_{n\in\mathbb{N}}$ nella classe compatta tale che $C_n \subset D_n$ e sia $\varepsilon > 0$ tale che $Q(\omega, D_n) \leq Q(\omega, C_n) + \varepsilon 2^{-n}$. Abbiamo che $\bigcap_n C_n \subset \bigcap_n D_n = \emptyset$. Per la proprietà della classe compatta esiste un $N \geq 1$ tale che $\bigcap_{n=1}^N C_n = \emptyset$. Sia ora $D_N = \bigcap_{n=1}^N D_n$. Abbiamo che $D_N \subset \bigcap_{n=1}^N D_n \setminus \bigcap_{n=1}^N C_n = \bigcup_{n=1}^N (D_n \setminus C_n)$. Per l'additività quindi $Q(\omega, D_N) \leq \sum_{n=1}^N (Q(\omega, D_n) - Q(\omega, C_n)) \leq \sum_{n=1}^\infty (Q(\omega, D_n) - Q(\omega, C_n)) \leq \varepsilon$. per l'arbitrarietà di ε ho che $Q(\omega, D_n) \to 0$.

Si estende poi all'unica misura di probabilità sulla σ-algebra $\mathcal{F}$.

Se $\omega \in \mathcal{N}$, sia $Q(\omega, A) = P(A)$. In questo modo abbiamo che $Q(\omega, \cdot)$ è una probabilità sullo spazio $(\Omega, \mathcal{F})$ e che $Q(\cdot, A)$ è misurabile ed è una versione della probabilità condizionata alla σ-algebra $\mathcal{G}$. $\square$

Vediamo ora in che casi queste ipotesi sono soddisfatte.

Lemma 3.2.3. *Sia E uno spazio metrico, P una misura finita su $(E, \mathcal{B}(E))$. Allora per ogni boreliano A e per ogni $\varepsilon > 0$ esistono un chiuso C ed un aperto G con $C \subset A \subset G$ e tale che*

$$P(G \setminus C) < \varepsilon. \tag{3.2}$$

Dimostrazione. Supponiamo che A sia chiuso e sia d la distanza in E. Allora se $G_\eta = \{x : d(x, A) < \eta\}$, G_η è aperto e $G_\eta \searrow A$ per $\eta \to 0$; quindi $P(G_\eta) \searrow P(A)$. Scegliendo $A = C$ e $G_\eta = G$ per η abbastanza piccolo ho la (3.2). Per concludere la dimostrazione basta far vedere che la classe degli A che verificano la tesi è una σ-algebra. Se A verifica la tesi anche A^c la verifica, infatti C^c è aperto e G^c è chiuso, $G^c \subset A^c \subset C^c$ e $P(C^c \setminus G^c) = P(G \setminus C) < \varepsilon$. Prendiamo ora $(A_n)_{n\in\mathbb{N}}$ una successione di boreliani per cui è soddisfatta la tesi e siamo C_n e G_n rispettivamente dei chiusi e degli aperti tali che $C_n \subset A_n \subset G_n$ e $P(G_n \setminus C_n) < \frac{\varepsilon}{2n+1}$. Sia $G = \bigcup_{n=1}^{\infty} G_n$ e $C = \bigcup_{n=1}^{\infty} C_n$, allora $P(G \setminus C) < \frac{\varepsilon}{2}$. Inoltre G è aperto e possiamo scegliere un k abbastanza grande per cui posto $C' = \bigcup_{n=1}^{k} C_n$, che è chiuso, si ha $P(C \setminus C') < \frac{\varepsilon}{2}$. Quindi abbiamo $C' \subset \bigcup_n A_n \subset G$ e $P(G \setminus C') < \varepsilon$. $\square$

Lemma 3.2.4. *Sia E uno spazio metrico completo e separabile, P una misura di probabilità su $(E, \mathcal{B}(E))$. Allora per ogni $\varepsilon > 0$ esiste un compatto K_ε tale che $P(K_\varepsilon) > 1 - \varepsilon$.*

Dimostrazione. sia $(x_n)_{n\in\mathbb{N}}$ una successione densa in E e per ogni i, n sia $B_{n,i}$ la sfera di centro x_i e raggio $\frac{1}{n}$. Allora per ogni n ho $\bigcup_i B_{n,i} = E$. Sia ora j_n abbastanza grande affinché posto $A_n = \bigcup_{i=1}^{j_n} B_{n,i}$ si abbia $P(A_n) > 1 - \frac{\varepsilon}{2^n}$. Sia $A = \bigcap_n A_n$. La sua chiusura $\overline{A}$ è un insieme chiuso totalmente limitato in uno spazio metrico completo; ciò implica che è compatto e $P(\overline{A}) > P(A) > 1 - \varepsilon$. □

Teorema 3.2.5. *Se Ω è uno spazio metrico, completo e separabile (spazio polacco) e $\mathcal{F}$ la σ-algebra boreliana, allora esiste sempre una versione regolare della probabilità condizionata su $\mathcal{F}$.*

Dimostrazione. combinando i due lemmi si vede che ogni legge di probabilità su $(\Omega, \mathcal{F})$ è regolare rispetto alla classe compatta A formata dai sottoinsiemi compatti di Ω. Inoltre poiché la topologia è a base numerabile si ha che $\mathcal{F}$ è numerabilmente generata. □

3.3 Esempi

$\Omega = \mathbb{R}^n$ questo è uno spazio di Banach separabile e applicando il teorema possiamo affermare che esiste sempre una versione regolare sulla σ-algebra dei boreliani.

$\Omega = C([0,1], \mathbb{R}^n)$ è uno spazio di Banach separabile con la metrica della convergenza uniforme. Applicando il teorema quindi possiamo dire che esiste sempre una versione regolare sulla σ-algebra dei boreliani.

Per dare un altro esempio interessante dimostriamo alcuni risultati preliminari.

Lemma 3.3.1. *Sia $X : (\Omega, \mathcal{F}, P) \to (E, \mathcal{B}(E))$ una v.a., Q la legge di X e supponiamo che Q sia regolare su $\mathcal{B}(E)$ rispetto a una classe compatta $\mathcal{K}$. Allora $\sigma(X)$ contiene una classe compatta rispetto alla quale P è regolare su $\sigma(X)$.*

Dimostrazione. Essendo Q legge di X ed essendo regolare su $\mathcal{B}(E)$ rispetto a $\mathcal{K}$ abbiamo $P(X^{-1}(B)) = \sup_{K \in \mathcal{K}, K \subset B} P^{-1}(K)$ per ogni $B \in \mathcal{B}(E)$. Siano K_n classe di elementi di $\mathcal{K}$ tale che $\bigcap_{n=1}^{\infty} K_n = \emptyset$. Consideriamo quindi $X^{-1}(K_n)$. Supponiamo che $\bigcap_{N=1}^{\infty} X^{-1}(K_n) = \emptyset$, ma $\bigcap_{N=1}^{\infty} X^{-1}(K_n) = X^{-1}(\bigcap_{n=1}^{\infty} K_n) = X^{-1}(\bigcap_{n=1}^{m} K_n) = \bigcap_{n=1}^{m} X^{-1}(K_n)$. Questo vuol dire che

$X^{-1}(\mathcal{K})$, che è contenuta in $\sigma(X)$, è la nostra classe compatta rispetto la quali P è regolare. $\square$

Lemma 3.3.2. *Sia $X : (\Omega, \mathcal{F}, P) \to (E, \mathcal{B}(E))$ una v.a. Allora se $\mathcal{B}(E)$ è numerabilmente generata anche $\sigma(X)$ lo è.*

Dimostrazione. Sia B' la classe numerabile che genera $\mathcal{B}(E)$. Allora $\sigma(X) = \{X^{-1}(H) : H \in \sigma(B')\}$. Consideriamo l'insieme numerabile $\Sigma = \{X^{-1}(H) : H \in B'\}$. Ovviamente $\sigma(\Sigma) \subset \sigma(X)$. Facciamo vedere l'implicazione inversa. $X^{-1}(H) \in \sigma(X)$ può essere o complementare o unione di $X^{-1}(H_n)$ con (H_n) elementi di B'.In questo caso abbiamo $X^{-1}(\bigcup_n H_n) = \bigcup_n X^{-1}(H_n)$ e questo appartiene a $\sigma(\Sigma)$. Se è complementare invece $X^{-1}(H_1^c) = (X^{-1}(H_1))^c \in \sigma(\Sigma)$.

Allora le due σ-algebra coincidono e quindi $\sigma(X)$ è numerabilmente generata. $\square$

Corollario 3.3.3. *Sia $(\Omega, \mathcal{F}, P)$ qualunque e X v.a. a valori in uno spazio metrico completo e separabile, allora su $\sigma(X)$ esiste una versione regolare della probabilità condizionata rispetto a una qualunque σ-algebra contenuta in $\mathcal{F}$.*

Bibliografia

[1] A.PASCUCCI, *stocastico per la finanza*, Springer, 2007

[2] A.PASCUCCI, *Note lezioni*, dispense del corso ‘probabilità e statistica 1’, Bologna, 2014

[3] F.FLANDOLI, *Note lezioni*, dispense del corso ‘Istituzioni di probabilità’, Pisa, 2014

[4] J.NEVEU, *Mathematical foundations of the calculus of probability*, Holden-Day, San Francisco, 1965

[5] P.BALDI, *Equazioni differenziali stocastiche e applicazioni*, Bologna, Pitagora Editrice, 1984

[6] P. BILLINGSLEY, *Probability and Measure*, Third edition, New York, John Wiley and Sons, 1995

[7] R. ASH, *Probability and measure theory*, Second edition, Academic Press, 1999