Appunti del corso Probabilità e Finanza
A.A. 2003/2004

Capitolo 3. Condizionamento

Paolo Baldi
Lucia Caramellino

http://www.mat.uniroma2.it/~caramell

Indice

Capitolo 3

Condizionamento

Sia $(\Omega, \mathscr{F}, \mathbb{P})$ uno spazio di probabilità. Dunque, ricordiamo,

Ω è un insieme (lo *spazio campionario*);

$\mathscr{F}$ è una σ-algebra di sottoinsiemi di Ω (gli *eventi*) ;

$\mathbb{P}$ è una probabilità su $(\Omega, \mathscr{F})$, cioè $\mathbb{P}$ è un'applicazione $\mathscr{F} \to [0,1]$ t.c.

$\mathbb{P}(\Omega) = 1$,

se $A_n \cap A_m = \emptyset$, $n \neq m$, $\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ (σ-*additività*).

Ricordiamo che, dati due eventi $A, B \in \mathscr{F}$ con $\mathbb{P}(B) > 0$, la probabilità condizionata di A dato B è definita da

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

Intuitivamente, $\mathbb{P}(A \,|\, B)$ rappresenta la probabilità del verificarsi dell'evento A, una volta noto il fatto che l'evento B si è verificato.
Vogliamo ora definire la *media condizionale* di una v.a. X data una σ-algebra $\mathscr{G} \subset \mathscr{F}$ e, in seguito, la *probabilità condizionale* di un evento qualsiasi $A \in \mathscr{F}$ data una σ-algebra $\mathscr{G}$. Poiché si tratta di oggetti alquanto delicati e spesso inizialmente di difficile comprensione, procederemo per gradi.

3.1 Media condizionale: il caso discreto

Cominciamo con un caso semplice, anche se tipico di condizionamento.

Esempio 3.1.1. Supponiamo che (X, Y) sia una coppia di v.a. discrete, con densità (discreta) congiunta p_{XY} e densità marginali p_X e p_Y. Siano E_X ed E_Y i possibili valori per X e Y rispettivamente, e supponiamo che $p_Y(y) > 0$ per ogni $y \in E_Y$. Per X, supponiamo che abbia speranza matematica finita.

Abbiamo visto che la densità discreta di X condizionata a $\{Y = y\}$, è data da
$$p_{X \mid Y}(x \mid y) = \frac{p_{XY}(x, y)}{p_Y(y)}, \quad \text{purché } y \in E_Y.$$
La media condizionale di X dato $\{Y = y\}$ è semplicemente la media di X quando per X si consideri la distribuzione condizionale:

$$\begin{aligned}
\mathbb{E}(X \mid Y = y) = & \textstyle\sum_{x \in E_X} x\, p_{X \mid Y}(x \mid y) = \frac{1}{p_Y(y)} \sum_{x \in E_X} x\, p_{X\,Y}(x, y) \\
= & \frac{1}{p_Y(y)} \sum_{(x,y) \in E_X \times \{y\}} x\, p_{X\,Y}(x, y) \\
= & \frac{1}{p_Y(y)}\, \mathbb{E}(X\, 1_{\{Y=y\}}).
\end{aligned}$$

Se poniamo, per $y \in E_Y$, $h_X(y) = \mathbb{E}(X\, 1_{\{Y=y\}})/p_Y(y)$, la v.a. $Z = h_X(Y)$ dà la media di X dato il risultato (aleatorio) Y. Ad essa può quindi essere dato il significato di *media condizionale di X dato Y*.
Osserviamo che $Z = h_X(Y)$ verifica le seguenti proprietà.

1. $Z = h_X(Y)$ è $\sigma(Y)$-misurabile. Infatti è una funzione (misurabile) di Y.

2. Per ogni $B \subset E_Y$, si ha
$$\mathbb{E}(h_X(Y)\, 1_{\{Y \in B\}}) = \mathbb{E}(X\, 1_{\{Y \in B\}})$$
Infatti,
$$\mathbb{E}(h_X(Y)\, 1_{\{Y \in B\}}) = \sum_{y \in B} h_X(y)\, p_Y(y) = \sum_{y \in B} \mathbb{E}(X\, 1_{\{Y=y\}}) = \mathbb{E}(X\, 1_{\{Y \in B\}}).$$

Infine, osserviamo che la v.a. $Z = h_X(Y)$ è l'unica v.a. che soddisfa alle condizioni 1. e 2. Infatti, $\sigma(Y)$ è la σ-algebra generata dalla partizione, al più numerabile, $\mathscr{C}_Y = \{Y^{-1}(\{y\})\,;\, y \in E_Y\}$. Sappiamo (Proposizione 2.1.18) che Y è costante sugli elementi della partizione. Dunque, se $\tilde{Z} = \tilde{h}(Y)$ è un'altra v.a. $\sigma(Y)$-misurabile, se essa soddisfa a 2. si deve avere, scegliendo $B = \{Y = y\} = Y^{-1}(y)$,
$$\mathbb{E}(X\, 1_{\{Y=y\}}) = \mathbb{E}(\tilde{h}(Y)\, 1_{\{Y=y\}}) = \tilde{h}(y)\, \mathbb{E}(1_{\{Y=y\}})$$
e cioè $\tilde{h}(y) = \mathbb{E}(X\, 1_{\{Y=y\}})/\mathbb{P}(Y = y) = h_X(y)$.

Analizziamo ora un caso un po' più generale.

Esempio 3.1.2. Sia X una v.a. con speranza matematica finita e sia $\mathscr{G}$ una σ-algebra, con $\mathscr{G} \subset \mathscr{F}$. Supponiamo che $\mathscr{G}$ sia generata da una partizione al più numerabile: $\mathscr{G} = \sigma(\mathscr{C})$, con $\mathscr{C} = \{C_i\}_{i \in I}$. Supponiamo anche che $\mathbb{P}(C_i) > 0$ per ogni $i \in I$.
Osserviamo che nell'Esempio 3.1.1 abbiamo implicitamente lavorato con questo tipo di ingredienti. Infatti,

$$\sigma(Y) = \mathscr{G} = \sigma(\mathscr{C}_Y), \quad C \in \mathscr{C}_Y \text{ se e solo se } C = Y^{-1}(\{y\}), \text{ con } y \in E_Y.$$

Ricordiamo inoltre che la media condizionale di X dato Y è

$$Z(\omega) = h_X(Y(\omega)) = \frac{\mathbb{E}(X\, 1_{\{Y=y\}})}{\mathbb{P}(Y=y)}, \quad \text{quando } \omega \in \{Y=y\}.$$

Ora, gli eventi $\{Y = y\}$, con $y \in E_Y$, sono gli atomi della partizione $\mathscr{C}_Y$ che genera $\mathscr{G} = \sigma(Y)$. Volendo quindi generalizzare ad una qualsiasi σ-algebra $\mathscr{G}$ generata da una partizione $\mathscr{C}$, viene naturale considerare la v.a.

$$Z(\omega) = \sum_{i \in I} \frac{\mathbb{E}(X\, 1_{C_i})}{\mathbb{P}(C_i)} 1_{\{\omega \in C_i\}} = \frac{\mathbb{E}(X\, 1_{C_k})}{\mathbb{P}(C_k)} \quad \text{quando } \omega \in C_k$$

Come vedremo in seguito, la v.a. Z è la media condizionale di X data la σ-algebra $\mathscr{G}$. Osserviamo che abbiamo fatto un paio di passi in avanti rispetto all'Esempio 3.1.1: abbiamo una σ-algebra $\mathscr{G}$ (almeno apparentemente) più generale e, soprattutto, non stiamo facendo ipotesi speciali sulla v.a. X: l'importante è che abbia senso la quantità $\mathbb{E}(X\, 1_{C_k})$, dunque l'importante è che X abbia media.
La v.a. Z verifica le seguenti proprietà, analoghe a quelle già viste nell'Esempio 3.1.1.

1. Z è $\mathscr{G}$-misurabile. Infatti, 1_{C_i} è $\mathscr{G}$-misurabile per ogni i, ed essendo Z una combinazione lineare delle funzioni 1_{C_i}, è $\mathscr{G}$-misurabile.

2. Per ogni $G \in \mathscr{G}$, si ha

$$\mathbb{E}(Z\, 1_G) = \mathbb{E}(X\, 1_G).$$

   Infatti, preso $G \in \mathscr{G}$, sia $I_G \subset I$ tale che $G = \cup_{i \in I_G} C_i$. Allora, $Z\, 1_G = \sum_{i \in I_G} \frac{\mathbb{E}(X\, 1_{C_i})}{\mathbb{P}(C_i)} 1_{C_i}$ e quindi

$$\begin{aligned} \mathbb{E}(Z\, 1_G) = & \sum_{i \in I_G} \frac{\mathbb{E}(X\, 1_{C_i})}{\mathbb{P}(C_i)} \mathbb{P}(C_i) = \sum_{i \in I_G} \mathbb{E}(X\, 1_{C_i}) \\ = & \ \mathbb{E}(X\, 1_{\cup_{i \in I_G} C_i}) = \mathbb{E}(X\, 1_G). \end{aligned}$$

Infine, osserviamo che anche in questo caso la v.a. Z è l'unica v.a. che soddisfa alle condizioni 1. e 2. Infatti, da 1. Z è $\mathscr{G}$-misurabile, dunque per la Proposizione 2.1.18 è costante sugli elementi della partizione: $Z(\omega) = z_i$ per ogni $\omega \in C_i$. Allora, da 2. si ha

$$\mathbb{E}(X\, 1_{C_i}) = \mathbb{E}(Z\, 1_{C_i}) = z_i\, \mathbb{E}(1_{C_i}) = z_i\, \mathbb{P}(C_i),$$

cioè $z_i = \mathbb{E}(X\, 1_{C_i})/\mathbb{P}(C_i)$, da cui la tesi.

3.2 Il caso generale

Negli Esempi 3.1.1 e 3.1.2, abbiamo visto che le proprietà 1. e 2. si rivelano caratteristiche della v.a. Z, che intuitivamente abbiamo chiamato *speranza condizionale di* X data la v.a. Y oppure data la σ-algebra $\mathscr{G}$. La definizione generale prende infatti in considerazione proprio queste due proprietà caratteristiche:

Definizione 3.2.1. Siano $(\Omega, \mathscr{F}, \mathbb{P})$ uno spazio di probabilità, X una v.a. avente speranza matematica finita e $\mathscr{G}$ una sotto σ-algebra di $\mathscr{F}$. Una versione $\mathbb{E}(X \mid \mathscr{G})$ della *media (o speranza) condizionale di* X *dato* $\mathscr{G}$ è una v.a. tale che:

(i) $\mathbb{E}(X \mid \mathscr{G})$ è $\mathscr{G}$-misurabile e integrabile;

(ii) per ogni $G \in \mathscr{G}$,
$$\mathbb{E}(\mathbb{E}(X \mid \mathscr{G})\, 1_G) = \mathbb{E}(X\, 1_G)$$

Qualora $\mathscr{G} = \sigma(Y)$, con Y v.a., allora $\mathbb{E}(X \mid \sigma(Y))$ prende il nome di speranza condizionale di X dato Y ed è denotata con il simbolo $\mathbb{E}(X \mid Y)$.

Abbiamo visto negli esempi che la speranza condizionale esiste se la σ-algebra $\mathscr{G}$ è generata da una partizione al più numerabile. Lasceremo a corsi più avanzati la dimostrazione dell'esistenza quando la σ-algebra è più generale. Del resto per in questo corso il caso di σ-algebre generate da una partizione sarà quello che interessa.
Nella Definizione 3.2.1 si parla di *versione* perché è abbastanza chiaro che non c'è unicità. Infatti, se Z è una v.a. $\mathscr{G}$-misurabile e avente speranza matematica finita e tale che $\mathbb{E}(1_G Z) = \mathbb{E}(1_G X)$ per ogni $G \in \mathscr{G}$ e Z' è un'altra v.a., sempre $\mathscr{G}$-misurabile, ma che differisce da Z solo su un evento di probabilità 0, allora si ha anche $\mathbb{E}(1_G Z') = \mathbb{E}(1_G X)$ per ogni $G \in \mathscr{G}$.

Quando $\mathscr{G} = \sigma(Y)$, il calcolo della speranza condizionale prende un aspetto particolare. Abbiamo infatti visto (Proposizione 2.1.20) che ogni v.a. $\sigma(Y)$-misurabile è della forma $h(Y)$, dove h è un'opportuna funzione misurabile. Quindi esiste una funzione misurabile Ψ_X tale che

$$\mathbb{E}(X \mid Y) = \Psi_X(Y).$$

Il calcolo della speranza condizionale, in questo caso, si riconduce quindi a quello della funzione Ψ_X. Con abuso di linguaggio, denoteremo

$$\Psi_X(y) = \mathbb{E}(X \mid Y = y),$$

ma attenzione a non confondere la notazione "$\mathbb{E}(X \mid Y = y)$" con la media condizionale di X dato l'evento $\{Y = y\}$ (che, osserviamo, in molti casi di interesse è un evento di probabilità nulla...). Nell'Esempio 3.1.1, abbiamo visto che

$$\Psi_X(y) = h_X(y) = \frac{\mathbb{E}(X\, 1_{\{Y=y\}})}{\mathbb{P}(Y = y)}.$$

In generale, la funzione Ψ_X resta determinata dalla relazione

$$\mathbb{E}(X 1_{\{Y\in B\}}) = \mathbb{E}(\Psi_X(Y) 1_{\{Y\in B\}}), \quad \text{per ogni boreliano } B \tag{3.1}$$

(ricordiamo infatti che $\sigma(Y) = \{Y^{-1}(B)\,;\, B \in \mathscr{B}(E)\}$). Osserviamo che, se X, Y sono v.a. a valori reali, allora siamo in grado di calcolare le due speranze matematiche che compaiono nella (3.1) non appena conosciamo le distribuzioni congiunte di X e Y. Vediamo, nel prossimo esempio, come questo fatto si può sfruttare per calcolare $\mathbb{E}(X \mid Y)$.

Esempio 3.2.2. Siano X e Y due v.a. reali congiuntamente assolutamente continue, con X avente speranza matematica finita. Indichiamo con f la densità di probabilità congiunta e con f_X e f_Y le due densità marginali. Supponiamo, ma solo per semplicità, che $f_Y(y) > 0$.
La speranza matematica a destra nella (3.1) vale dunque

$$\mathbb{E}(X 1_{\{Y\in B\}}) = \int_B dy \int_{\mathbb{R}} x f(x, y)\, dx$$

mentre

$$\mathbb{E}(\Psi_X(Y) 1_{\{Y\in B\}}) = \int_B \Psi_X(y) f_Y(y)\, dy$$

Queste due quantità risultano uguali per ogni scelta del boreliano B se scegliamo

$$\Psi_X(y) = \frac{1}{f_Y(y)} \int_{\mathbb{R}} x f(x, y)\, dx = \int_{\mathbb{R}} x f_{X\mid Y}(x \mid y)\, dx$$

dove

$$f_{X\mid Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

è la (vecchia!) densità condizionale. Per concludere dobbiamo solo verificare che la v.a. $\Psi_X(Y)$ ha speranza matematica finita. Infatti

$$\int_{\mathbb{R}} |\Psi_X(y)|\, f_Y(y)\, dy \le \int_{\mathbb{R}} dx \int_{\mathbb{R}} |x|\, f(x, y)\, dy = \int_{\mathbb{R}} |x|\, f_X(x)\, dx < \infty$$

poiché supponiamo che X abbia speranza matematica finita.
In conclusione, la media di X condizionata a Y è la media di X fatta sotto la densità condizionale $f_{X\mid Y}(x \mid y)$, sostituendo poi ad y il suo valore generico (aleatorio) Y: esattamente come vorrebbe l'intuizione!

3.3 Proprietà della speranza condizionale

Vediamo ora alcune proprietà delle speranze condizionali. Avvertenza: nel seguito l'abbreviazione "q.c." sta per "quasi certamente". Significa che le proprietà in questione (uguaglianze, disuguaglianze etc.) sono verificate per $\omega \in \Omega \setminus \mathcal{N}$, dove $\mathcal{N}$ denota un evento tale che $\mathbb{P}(\mathcal{N}) = 0$.

1. Se $X = c$, allora $\mathbb{E}(X \,|\, \mathscr{G}) = c$ q.c.: immediato!

2. $\mathbb{E}(\mathbb{E}(X \,|\, \mathscr{G})) = \mathbb{E}(X)$. Infatti basta prendere $G = \Omega \in \mathscr{G}$ nel punto (ii) della Definizione 3.2.1.

3. (Linearità) Se X e Y hanno media allora si ha $\mathbb{E}(\alpha X + \beta Y \,|\, \mathscr{G}) = \alpha\mathbb{E}(X \,|\, \mathscr{G}) + \beta\mathbb{E}(Y \,|\, \mathscr{G})$ q.c., per ogni $\alpha, \beta \in \mathbb{R}$. Infatti, per ipotesi $\mathbb{E}(X \,|\, \mathscr{G})$ e $\mathbb{E}(Y \,|\, \mathscr{G})$ sono $\mathscr{G}$-misurabili ed hanno speranza matematica finita. Quindi lo stesso vale per $Z := \alpha\mathbb{E}(X \,|\, \mathscr{G}) + \beta\mathbb{E}(Y \,|\, \mathscr{G})$. Inoltre, se $G \in \mathscr{G}$,

$$\begin{aligned}\mathbb{E}(Z\, 1_G) &= \alpha\mathbb{E}(\mathbb{E}(X \,|\, \mathscr{G})\, 1_G) + \beta\mathbb{E}(\mathbb{E}(Y \,|\, \mathscr{G})\, 1_G)\\ &= \alpha\mathbb{E}(X\, 1_G) + \beta\mathbb{E}(Y\, 1_G) = \mathbb{E}((\alpha X + \beta Y)\, 1_G),\end{aligned}$$

dunque $Z := \alpha\mathbb{E}(X \,|\, \mathscr{G}) + \beta\mathbb{E}(Y \,|\, \mathscr{G})$ è una versione della speranza condizionale di $\alpha X + \beta Y$ dato $\mathscr{G}$.

4. (Monotonia) Se $\mathbb{P}(X \geq Y) = 1$ allora $\mathbb{E}(X \,|\, \mathscr{G}) \geq \mathbb{E}(Y \,|\, \mathscr{G})$ q.c.
Infatti, poniamo $G = \{\omega\,;\, \mathbb{E}(X \,|\, \mathscr{G}) < \mathbb{E}(Y \,|\, \mathscr{G})\}$. Evidentemente $G \in \mathscr{G}$. Se fosse $\mathbb{P}(G) > 0$ allora si avrebbe $0 > \mathbb{E}((\mathbb{E}(X \,|\, \mathscr{G}) - \mathbb{E}(Y \,|\, \mathscr{G}))\, 1_G)$, mentre invece si ha

$$\mathbb{E}((\mathbb{E}(X \,|\, \mathscr{G}) - \mathbb{E}(Y \,|\, \mathscr{G}))\, 1_G) = \mathbb{E}(\mathbb{E}(X - Y \,|\, \mathscr{G})\, 1_G) = \mathbb{E}((X - Y)\, 1_G) \geq 0,$$

il che è assurdo.

5. (Disuguaglianza di Jensen) Se Φ è una funzione continua convessa e tale che $\Phi(X)$ abbia speranza matematica finita, allora

$$\mathbb{E}(\Phi(X) \,|\, \mathscr{G}) \geq \Phi(\mathbb{E}(X \,|\, \mathscr{G})) \quad \text{q.c.}$$

In particolare, se $p \geq 1$ e $\mathbb{E}(|X|^p) < +\infty$, allora possiamo scegliere $\Phi(x) = |x|^p$ e dunque $\mathbb{E}(|X|^p \,|\, \mathscr{G}) \geq |\mathbb{E}(X \,|\, \mathscr{G})|^p$, q.c.
Infatti, se Φ è convessa allora si può dimostrare che $\Phi(x) = \sup_{f \in A_\Phi} f(x)$, dove A_Φ è l'insieme delle funzioni lineari affini f che minorano Φ: $A_\Phi = \{f\,;\, f(x) = ax + b \text{ e } f(x) \leq \Phi(x)\}$ (vedi figura). Allora,

$$\mathbb{E}(\Phi(X) \,|\, \mathscr{G}) \geq \mathbb{E}(f(X) \,|\, \mathscr{G}) = f(\mathbb{E}(X \,|\, \mathscr{G})),$$

da cui segue che

$$\mathbb{E}(\Phi(X) \,|\, \mathscr{G}) \geq \sup_{f \in A_\Phi} f(\mathbb{E}(X \,|\, \mathscr{G})) = \Phi(\mathbb{E}(X \,|\, \mathscr{G})).$$

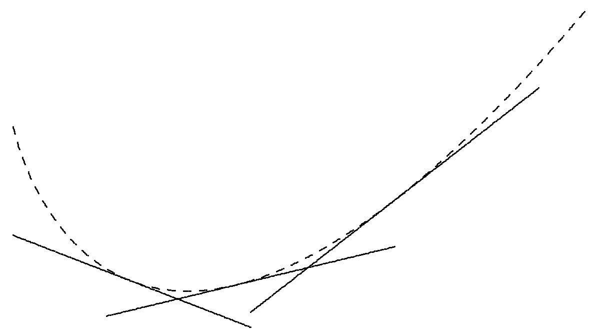

Figura 3.1 Una funzione convessa e continua è sempre inviluppo superiore delle funzioni lineari affini che la minorano.

6. Se Y è $\mathscr{G}$-misurabile allora $\mathbb{E}(XY \mid \mathscr{G}) = Y\,\mathbb{E}(X \mid \mathscr{G})$, q.c. In particolare, $\mathbb{E}(Y \mid \mathscr{G}) = Y$, q.c.
Faremo la dimostrazione solo nel caso in cui Y è prende un numero finito di valori, $y_1, \dots, y_n$: $Y = \sum_{i \le n} y_i\, 1_{\{Y=y_i\}}$. Se poniamo $G_i = \{Y = y_i\}$, allora $G_i \in \mathscr{G}$ perché Y è $\mathscr{G}$-misurabile. Ora, preso $G \in \mathscr{G}$, si ha:

$$\begin{aligned}\mathbb{E}(XY\,1_G) &= \sum_{i \le n} y_i \mathbb{E}(X\,1_{G_i \cap G}) = \sum_{i \le n} y_i \mathbb{E}(\mathbb{E}(X \mid \mathscr{G})\,1_{G_i \cap G}) \\ &= \mathbb{E}\Big(\sum_{i \le n} y_i\,1_{G_i}\,\mathbb{E}(X \mid \mathscr{G})\,1_G\Big) = \mathbb{E}(Y\,\mathbb{E}(X \mid \mathscr{G})\,1_G).\end{aligned}$$

7. Siano $\mathscr{A}, \mathscr{G} \subset \mathscr{F}$ due σ-algebre tali che $\mathscr{A} \subset \mathscr{G}$. Allora: $\mathbb{E}(X \mid \mathscr{A}) = \mathbb{E}(\mathbb{E}(X \mid \mathscr{A}) \mid \mathscr{G}) = \mathbb{E}(\mathbb{E}(X \mid \mathscr{G}) \mid \mathscr{A})$, q.c.
$\mathbb{E}(X \mid \mathscr{A})$ è $\mathscr{A}$-misurabile e quindi è $\mathscr{G}$-misurabile perché $\mathscr{A} \subset \mathscr{G}$. Dunque, per il punto precedente, $\mathbb{E}(X \mid \mathscr{A}) = \mathbb{E}(\mathbb{E}(X \mid \mathscr{A}) \mid \mathscr{G})$. Dimostriamo ora che $\mathbb{E}(X \mid \mathscr{A}) = \mathbb{E}(\mathbb{E}(X \mid \mathscr{G}) \mid \mathscr{A})$: preso $A \in \mathscr{A}$, allora A appartiene anche a $\mathscr{G}$ e

$$\mathbb{E}(\mathbb{E}(X \mid \mathscr{G})\,1_A) = \mathbb{E}(X\,1_A).$$

8. Sia X indipendente da G. Allora, $\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(X)$, q.c.
Infatti, se $G \in \mathscr{G}$ allora X e $Y = 1_G$ sono indipendenti, quindi

$$\mathbb{E}(X\,1_G) = \mathbb{E}(X)\mathbb{E}(1_G) = \mathbb{E}(\mathbb{E}(X)\,1_G).$$

9. Ora sappiamo che la speranza condizionale $\mathbb{E}(X \mid \mathscr{G})$ si calcola facilmente se X è indipendente da $\mathscr{G}$ (vedi 8.) oppure se X è già $\mathscr{G}$-misurabile. C'è un'altra situazione, molto frequente e che combina queste due situazioni, in cui il calcolo è abbastanza facile. Siano X una v.a. $\mathscr{G}$-misurabile e Y una v.a. indipendente da $\mathscr{G}$. Per ogni f misurabile tale che $f(X,Y)$ ha speranza finita si ha $\mathbb{E}(f(X,Y) \mid \mathscr{G}) = \bar{f}(X)$ q.c., dove $\bar{f}(x) = \mathbb{E}(f(x,Y))$.
Lo dimostreremo solo nel caso in cui la funzione f è della forma $f(x,y) = f_1(x)f_2(y)$. Il caso generale si dimostra usando il fatto che ogni funzione

$f(x,y)$ si può approssimare con delle combinazioni lineari di funzioni del tipo $f_1(x)f_2(y)$ e ne lasceremo la dimostrazione ad un corso più avanzato. Se f è della forma indicata, allora $\bar{f}(x) = f_1(x)\mathbb{E}(f_2(Y))$. Usando 6. e 8.,

$$\begin{aligned}\mathbb{E}(f(X,Y)\,|\,\mathscr{G}) = \mathbb{E}(f_1(X)\,f_2(Y)\,|\,\mathscr{G}) = f_1(X)\,\mathbb{E}(f_2(Y)\,|\,\mathscr{G}) = \\ = f_1(X)\,\mathbb{E}(f_2(Y)) = \bar{f}(X).\end{aligned}$$

10. Sia X una v.a. con momento secondo finito. Grazie alla disuguaglianza di Jensen, sappiamo che anche $\mathbb{E}(X\,|\,\mathscr{G})$ ha momento secondo finito (si scelga $p = 2$ nel punto 5.). . Mostriamo ora una delle proprietà più importanti della speranza condizionale: si ha, per ogni v.a. Z che sia $\mathscr{G}$-misurabile e di quadrato integrabile,

$$\mathbb{E}[(X - \mathbb{E}(X\,|\,\mathscr{G}))^2] \le \mathbb{E}[(X - Z)^2] \tag{3.2}$$

e, per di più la disuguaglianza precedente è stretta a meno che non sia $Z = \mathbb{E}(X\,|\,\mathscr{G})$ q.c. In altre parole $\mathbb{E}(X\,|\,\mathscr{G})$ è la v.a. $\mathscr{G}$-misurabile che meglio approssima X nel senso dei minimi quadrati.
Si noti l'analogia con $\mathbb{E}(X)$: sia $\mathbb{E}(X)$ che $\mathbb{E}(X\,|\,\mathscr{G})$ sono le migliori stime di X nel senso dei minimi quadrati, nel primo caso nella classe delle costanti mentre nel secondo caso nello spazio delle v.a. con momento secondo finito e $\mathscr{G}$-misurabili.
Dimostriamo la (3.2). Presa una qualunque v.a. Z $\mathscr{G}$-misurabile e di quadrato integrabile, si ha

$$\begin{aligned}\mathbb{E}[(X - Z)^2] = \;& \mathbb{E}[(X - \mathbb{E}(X\,|\,\mathscr{G}))^2] + \mathbb{E}[(Z - \mathbb{E}(X\,|\,\mathscr{G}))^2] \\ & -2\mathbb{E}[(X - \mathbb{E}(X\,|\,\mathscr{G}))(Z - \mathbb{E}(X\,|\,\mathscr{G}))].\end{aligned}$$

Ma $Z - \mathbb{E}(X\,|\,\mathscr{G})$ è una v.a. $\mathscr{G}$-misurabile e dunque

$$\begin{aligned}& \mathbb{E}[(X - \mathbb{E}(X\,|\,\mathscr{G}))(Z - \mathbb{E}(X\,|\,\mathscr{G}))] \\ &= \mathbb{E}\big[\mathbb{E}[(X - \mathbb{E}(X\,|\,\mathscr{G}))(Z - \mathbb{E}(X\,|\,\mathscr{G}))]\,\big|\,\mathscr{G}\big] \\ &= (Z - \mathbb{E}(X\,|\,\mathscr{G}))\mathbb{E}\big[\mathbb{E}[X - \mathbb{E}(X\,|\,\mathscr{G})\,\big|\,\mathscr{G}]\big] = 0.\end{aligned}$$

Dunque,

$$\mathbb{E}[(X - Z)^2] \ge \mathbb{E}[(X - \mathbb{E}(X\,|\,\mathscr{G}))^2]$$

e si ha uguaglianza se e solo se $Z = \mathbb{E}(X\,|\,\mathscr{G})$, da cui la tesi.

3.4 Probabilità condizionale

Definizione 3.4.1. Sia $A \in \mathscr{F}$ e sia $\mathscr{G}$ una una sotto-σ-algebra di $\mathscr{F}$. Una versione della *probabilità condizionale* di A dato $\mathscr{G}$ è data da

$$\mathbb{P}(A\,|\,\mathscr{G}) = \mathbb{E}(1_A\,|\,\mathscr{G}).$$

Così come nel caso della media condizionale, quando $\mathscr{G} = \sigma(Y)$ useremo la notazione $\mathbb{P}(A\,|\,Y)$.

Ricordiamo che, anche in questo caso, occorre parlare di "versione": analogamente alla media condizionale, la probabilità condizionale è definita a meno di insiemi di probabilità nulla, ovvero è definita q.c.

Esempio 3.4.2. Riprendendo gli esempi 3.1.1, 3.1.2 e 3.2.2, possiamo scrivere:

- se (X, Y) è una coppia di v.a. discrete, di densità discreta congiunta $p_{X\,Y}$ e densità (discrete) marginali p_X e p_Y allora q.c.

$$\mathbb{P}(X \in B \,|\, Y) = h_B(Y), \quad \text{dove } h_B(y) = \mathbb{P}(X \in B \,|\, Y = y) = \sum_{x \in B} \frac{p_{X\,Y}(x, y)}{p_Y(y)}$$

  per ogni $B \subset E_X$ (abbiamo scelto in particolare $A = \{X \in B\}$);

- se $\mathscr{G} = \sigma(\mathscr{C})$, con $\mathscr{C} = \{C_i\}_{i \in I}$ partizione al più numerabile di Ω, e se $A \in \mathscr{F}$ allora q.c.

$$\mathbb{P}(A \,|\, \mathscr{G}) = \sum_i \frac{\mathbb{P}(A \cap C_i)}{\mathbb{P}(C_i)} 1_{C_i};$$

- se (X, Y) è una coppia di v.a. assolutamente continue, di densità continua congiunta $p_{X\,Y}$ e densità (continue) marginali p_X e p_Y allora q.c.

$$\mathbb{P}(X \in B \,|\, Y) = \Psi_B(Y), \quad \text{dove } \Psi_B(y) = \int_B \frac{p_{X\,Y}(x, y)}{p_Y(y)} \, dy,$$

  per ogni $B \in \mathscr{E}$ (abbiamo scelto in particolare $A = \{X \in B\}$).

La probabilità condizionale verifica le due proprietà basilari della probabilità:

(i) per ogni $A \in \mathscr{F}$, $\mathbb{P}(A^c \,|\, \mathscr{G}) = 1 - \mathbb{P}(A \,|\, \mathscr{G})$, q.c.: immediata conseguenza del fatto che $1 = 1_A + 1_{A^c}$.

(ii) per ogni $A_1, A_2, \ldots \in \mathscr{F}$ tali che $A_n \cap A_m = \emptyset$ per $n \neq m$ allora $\mathbb{P}(\cup_n A_n \,|\, \mathscr{G}) = \sum_n \mathbb{P}(A_n \,|\, G)$, q.c.

Lo dimostriamo solo nel caso in cui $\mathscr{G}$ è generata da una partizione $\mathscr{C} = \{C_i\}_{i \in I}$ al più numerabile. In tal caso, ponendo $I_{\mathcal{N}} = \{i \in I : \mathbb{P}(C_i = 0)\}$, allora ovviamente $\mathbb{P}(\cup_{i \in I_{\mathcal{N}}} C_i) = 0$ ed è quindi immediato vedere che

$$\mathbb{E}(X \,|\, \mathscr{G}) = \sum_{i \notin I_{\mathcal{N}}} \frac{\mathbb{E}(X\, 1_{C_i})}{\mathbb{P}(C_i)} 1_{C_i}, \quad \text{q.c.}$$

Ma allora, q.c. si hanno le seguenti uguaglianze:

$$\begin{aligned}\mathbb{P}(\cup_n A_n \,|\, \mathscr{G})(\omega) &= \sum_{i \notin I_{\mathcal{N}}} \frac{\mathbb{P}((\cup_n A_n) \cap C_i)}{\mathbb{P}(C_i)} 1_{C_i} = \sum_{i \notin I_{\mathcal{N}}} \sum_n \frac{\mathbb{P}(A_n \cap C_i)}{\mathbb{P}(C_i)} 1_{C_i} \\ &= \sum_n \sum_{i \notin I_{\mathcal{N}}} \frac{\mathbb{P}(A_n \cap C_i)}{\mathbb{P}(C_i)} 1_{C_i} = \sum_n \mathbb{P}(A_n \,|\, \mathscr{G}),\end{aligned}$$

da cui la tesi.

Si potrebbe allora pensare di poter costruire una v.a.

$$\mathbb{P}(\cdot,\cdot) \,:\, \mathscr{F} \times \Omega \longrightarrow \mathbb{R}$$

tale che

1. fissato $A \in \mathscr{F}$, $\omega \mapsto \mathbb{P}(A,\omega)$ è una versione di $\mathbb{P}(A \,|\, \mathscr{G})(\omega)$;
2. fissato $\omega \in \Omega$, $A \mapsto \mathbb{P}(A,\omega)$ è una probabilità su $(\Omega, \mathscr{F})$.

Qualora sia possibile, $\mathbb{P}(A,\omega)$ prende il nome di *versione regolare della probabilità condizionata*. In generale però, tale funzione $\mathbb{P}(A,\omega)$ non si può costruire. Infatti, abbiamo visto che, fissato $A \in \mathscr{F}$, $\mathbb{P}(A \,|\, \mathscr{G})$ è definita q.c.: esiste $\mathcal{N}_A \in \mathscr{F}$ tale che $\mathbb{P}(\mathcal{N}_A) = 0$ e $\mathbb{P}(A \,|\, \mathscr{F})(\omega)$ è ben definita per $\omega \notin \mathcal{N}_A$. Ne segue che, in generale,

$$\omega \mapsto \mathbb{P}(A \,|\, \mathscr{F})(\omega) \text{ è ben definita per } \omega \notin \mathcal{N} = \cup_{A \in \mathscr{F}} \mathcal{N}_A.$$

Ora, perché il tutto abbia senso occorre che $\mathcal{N}$ sia "trascurabile", cioè $\mathbb{P}(\mathcal{N}) = 0$. Ma, poiché $\mathscr{F}$ non è in generale numerabile e dunque $\mathcal{N} = \cup_{A \in \mathscr{F}} \mathcal{N}_A$ non è in generale un'unione numerabile di eventi,

- non è detto che $\mathcal{N}$ sia un evento, ovvero $\mathcal{N} \in \mathscr{F}$;
- se anche $\mathcal{N} \in \mathscr{F}$, non è detto che $\mathbb{P}(\mathcal{N}) = 0$.

Osserviamo infine che, poiché la probabilità condizionale di A è data dalla media condizionale della v.a. $X = 1_A$, molte delle proprietà viste per le medie condizionali valgono anche in questo caso. Ad esempio, se $A \in \mathscr{G}$ allora ovviamente $\mathbb{P}(A \,|\, \mathscr{G}) = 1_A$ q.c., così come se A è indipendente da $\mathscr{G}$ allora $\mathbb{P}(A \,|\, \mathscr{G}) = \mathbb{P}(A)$, q.c.

Bibliografia

[1] P. Baldi (2000) *Equazioni differenziali stocastiche e applicazioni.* Pitagora Editrice.

[2] P. Billingsley (1989) *Probability and measure.* Whiley.

[3] D. Williams (1991) *Probability with martingales.* Cambridge University Press.