UNIVERSITÀ DEGLI STUDI DI PADOVA
FACOLTÀ DI SCIENZE MM.FF.NN
CORSO DI LAUREA IN MATEMATICA

Elaborato di laurea

VALORE ATTESO CONDIZIONATO E SUE APPLICAZIONI

Relatore: prof. Marco Ferrante

Laureanda: Donata Melotto

Anno Accademico 2003-2004

Indice

Capitolo 1

Introduzione

L'oggetto di studio di questo elaborato è il valore atteso condizionato, trattato nel terzo e nel quinto capitolo.

Nel terzo capitolo lo si definisce negli spazi di probabilità discreti. Le sue applicazioni sono mostrate tramite vari esempi tra cui il problema del collezionista di figurine (il cosidetto "coupon collecting problem"). Si è dato rilievo a questo esempio perché si presenta una identità combinatoria che per i casi più semplici è dimostrabile algebricamente. Sempre nel terzo capitolo si definisce la varianza condizionata e si mostrano alcune sue applicazioni.

Nel quinto capitolo si definisce il valore atteso condizionato negli spazi di probabilità generali, rispetto ad un evento, ad una σ-algebra e ad una variabile aleatoria. È proposto un esempio di applicazione.

Si è ritenuto utile riassumere nel secondo capitolo i concetti di base della probabilità a partire dallo spazio campionario per terminare con il valore atteso, la varianza e la covarianza di una variabile aleatoria discreta. Per parlare di spazi di probabilità generali occorre sapere cosa sono le σ-algebre. Nel capitolo quarto si introduce il concetto di σ-algebra, misura di probabilità e di variabile aleatoria in modo sufficiente da potere trattare nel capitolo successivo il valore atteso condizionato nel caso generale.

Infine si riassume brevemente qui di seguito la storia della teoria della probabilità.

La probabilità nasce con il gioco d'azzardo: sembra che gli uomini vi si dedicassero già intorno al 3500 a.C. Pitture nelle tombe egiziane, risalenti a questo periodo, testimoniano che un gioco da tavolo, che si serviva di una specie di dado, era diffuso in Egitto. Si tratta dell'astragalo, un dado a quattro facce (inizialmente asimmetrico), ricavato dall'omonimo osso della capra o del montone. Dagli antichi autori greci e romani si sa che l'astragalo era usato in differenti tipi di gioco. L'uso dell'astragalo scomparve quasi del tutto intorno al 1600 d.C., per essere sostituito dal dado. Il primo dado a sei facce risale anch'esso al 3500 a.C. Nel periodo medievale gli europei giocavano d'azzardo con l'astragalo, il dado e con vari giochi da tavolo. Alla fine del secolo XV emersero le prime nozioni di probabilità. Il primo problema di probabilità effettivamente scritto, si trova in un libro del toscano Pacioli. Il problema riguarda la ripartizione della posta tra due giocatori in un gioco che viene interrotto. Questo problema fu risolto correttamente da B. Pascal nel 1654. Intorno al 1550 Gerolamo Cardano scrisse un libro intitolato *Liber de Ludo Alea*, in cui sviluppò la teoria della probabilità. Egli non pubblicò mai il suo libro, che fu scoperto solo a cento anni dalla sua morte. Ma in quel tempo la teoria della probabilità fu riscoperta da B. Pascal, P. Fermat e C. Huyghens. Pascal e Fermat gettarono le fondamenta in una serie di lettere che i due si inviarono tra il 1654 e il 1660. In queste lettere si riferivano schemi per risolvere problemi di probabilità di vario tipo. Sembra che Pascal sia stato spinto ad interessarsi a questo ramo della matematica da un noto giocatore d'azzardo, il cavaliere di Méré. Infine il giovane danese C. Huyghens scrisse un libro, pubblicato in latino nel 1656, intitolato *De Ratiociniis in Alea Ludo*. Questo testo è considerato il vero inizio della teoria della probabilità. Successivamente altri studiosi del secolo XIX si occuparono di probabilità: J. Bernoulli pubblicò un'opera in cui si trova la legge dei grandi numeri; D. Bernoulli, Lagrange, Laplace e Gauss applicarono i metodi del calcolo integrale; Bayes si occupò della probabilità a priori (che deriva dall'assunzione che certi eventi sono equiprobabili). Infine nella prima metà del secolo XX il russo Kolmogorov sviluppò una teoria assiomatica della probabilità.

Ci sono state essenzialmente tre vie per interpretare il concetto di probabilità:

- *probabilità a priori:* è la probabilità a cui tutte le persone possono convenire. Ad esempio tutti sono d'accordo nell'affermare che quando si lancia un dado non truccato, ogni numero ha probabilità di uscire pari a $1/6$. La probabilità a priori deriva da un'assunzione di uniformità, come detto sopra, ed è spesso basata su distribuzioni uniformi;

- *probabilità frequentista:* la probabilità di un evento è approssimativamente uguale alla frequenza calcolata su un certo numero di prove indipendenti. Ad esempio l'affermazione che nel lanciare una moneta si ha probabilità che esca testa pari a $1/2$, si interpreta così: se lanci una moneta cento volte, allora otterrai approssimativamente cinquanta volte testa e cinquanta volte croce;

- *probabilità soggettivista:* la probabilità di un evento è il grado di fiducia che un individuo attribuisce, in base alle proprie opinioni e alle informazioni di cui dispone, al verificarsi di un evento. Ad esempio prima dell'incidente di Chernobyl, la probabilità che si verificasse un incidente in un impianto nucleare era ritenuta insignificante. Dopo l'incidente la probabilità aumentò drasticamente.

Capitolo 2

Spazi di probabilità discreti

In questo capitolo vengono ricordati definizioni e teoremi necessari per approfondire l'argomento del valore atteso condizionato, oggetto del prossimo capitolo. Si definisce lo spazio di probabilità discreto, la probabilità, la variabile aleatoria discreta e il valore atteso.

2.1 Spazio campionario ed eventi

Supponiamo di volere fare un esperimento il cui esito non sia determinabile a priori. Supponiamo però di conoscere l'insieme di tutti gli esiti possibili di tale esperimento. Chiamiamo **spazio campionario** e lo indichiamo con $\mathbf{\Omega}$ l'insieme di tutti gli esiti possibili di un esperimento. Chiamiamo **evento** ogni sottoinsieme di Ω.

Dati gli eventi $E_1, E_2,...E_k$, $k \in \mathbb{N}$, $k \geq 2$, di uno spazio campionario Ω, consideriamo due eventi speciali di Ω: l'unione di questi eventi e l'intersezione di questi eventi. L'unione $\bigcup_{n=1}^{k} E_n$ consiste di tutti gli esiti che sono in E_n per almeno un valore di $n = 1, 2, ..., k$. L'intersezione $\bigcap_{n=1}^{k} E_n$ consiste di quegli esiti che sono in tutti gli eventi $E_1, E_2,...E_k$.

Esempio 2.1.1. Nell'esperimento del lancio di un dado, l'insieme degli esiti possibili è $\Omega = \{1, 2, 3, 4, 5, 6\}$ e l'evento "Esce il numero 1 nel lancio del dado" è $E_1 = \{1\}$ mentre l'evento "Esce un numero pari nel lancio del dado" è $E_2 = \{2, 4, 6\}$. Consideriamo anche $E_3 = \{1, 2, 6\}$.

Si ha $E_2 \cup E_3 = \{1, 2, 4, 6\}$ e $E_2 \cap E_3 = \{2, 6\}$.

2.2 Probabilità sugli eventi

Sia Ω un insieme finito o numerabile e sia $\mathcal{P}(\Omega)$ la famiglia dei sottoinsiemi di Ω (l'insieme delle parti di Ω). Una funzione $P : \mathcal{P}(\Omega) \longmapsto [0,1]$ si dice **probabilità** se soddisfa alle seguenti condizioni:

- $P(\Omega) = 1$;
- per ogni successione $(E_n)_{n\in\mathbb{N}}$ di sottoinsiemi di Ω (cioè eventi) a due a due disgiunti si ha
$$P(\bigcup_{n=1}^{+\infty} E_n) = \sum_{n=1}^{+\infty} P(E_n).$$

Chiamiamo **spazio di probabilità discreto** la coppia (Ω, P). L'ultima condizione è nota come proprietà di σ-additività.

Vediamo le proprietà di uno spazio di probabilità discreto (Ω, P) che sono conseguenza della definizione di probabilità, senza dimostrarle [1] :

1. $P(\varnothing) = 0$;
2. se $k \geq 2$ e $E_1, E_2, ...E_k$ sono eventi di Ω a due a due disgiunti, allora $P(\bigcup\limits_{n=1}^{k} E_n) = \sum\limits_{n=1}^{k} P(E_n)$;
3. se $\overline{E}$ è l'insieme complementare di E, allora $P(\overline{E}) = 1 - P(E)$;
4. se $E, F \subseteq \Omega$ e $E \subseteq F$ allora $P(F \backslash E) = P(F) - P(E)$;
5. se $E, F \subseteq \Omega$, allora $P(E \cup F) = P(E) + P(F) - P(E \cap F)$;
6. se $n \geq 2$ e $E_1, E_2, ...E_n$ sono eventi di Ω, allora $P(\bigcup\limits_{i=1}^{n} E_i) = \sum\limits_{i} P(E_i) - \sum\limits_{i<j} P(E_i \cap E_j) + \sum\limits_{i<j<k} P(E_i \cap E_j \cap E_k) - \sum\limits_{i<j<k<l} P(E_i \cap E_j \cap E_k \cap E_l) + ... + (-1)^{n+1} P(E_1 \cap E_2 ... \cap E_n)$;
7. *(continuità dal basso)* se $(A_n)_{n\in\mathbb{N}}$ è una successione crescente di eventi di Ω, cioè $A_n \subseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, allora $P(\bigcup\limits_{n=1}^{+\infty} A_n) = \lim\limits_{n\to+\infty} P(A_n)$;

[1] Per le prime tre e la quinta vedi **S.M. Ross**, *A First Course in Probability, 6th Edition*, Prentice Hall, Upper Saddle River (USA), 2002, pagg.30 e 32; per la quarta vedi **G.R. Grimmett, D. Stirzaker**, *Probability and Random Processes*, Clarendon Press, Oxford, 1982, pag.6; per la sesta vedi **Y.A. Rozanov**, *Probability Theory: a concise course*, Dover Publication, New York, 1977, pag.18. Per le ultime due si rimanda al caso generale (proposizione 4.1.1).

8. *(continuità dall'alto)* se $(A_n)_{n\in\mathbb{N}}$ è una successione decrescente di eventi di Ω, cioè $A_n \supseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, allora $P(\bigcap_{n=1}^{+\infty} A_n) = \lim_{n\to+\infty} P(A_n)$.

Esempio 2.2.1. Nell'esperimento del lancio di un dado gli eventi $\{1\}$,$\{2\}$,...,$\{6\}$ sono disgiunti, e se il dado non è truccato, abbiamo

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}.$$

Inoltre

$$P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}$$

e

$$P(\{2,4,6\} \cup \{1,2,6\}) = P(\{2,4,6\}) + P(\{1,2,6\}) - P(\{2,6\}) =$$
$$= \frac{1}{2} + \frac{1}{2} - \frac{1}{3} = \frac{2}{3}.$$

2.3 Condizionamento e indipendenza

Dato uno spazio di probabilità discreto (Ω, P) e due eventi $E, F \subseteq \Omega$ tali che $P(F) > 0$, chiamiamo **probabilità condizionata** di E rispetto a F, ossia la probabilità che si verifichi E sapendo che si è verificato F, la seguente quantità:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(EF)}{P(F)}.$$

La funzione $P(\cdot|F) : \mathcal{P}(\Omega) \longmapsto [0,1]$ gode delle proprietà tipiche di una probabilità [2]:

1. $P(\Omega|F) = 1$;
2. se $(E_n)_{n\in\mathbb{N}}$ è una successione di eventi di Ω disgiunti a due a due, allora $P(\bigcup_{n=1}^{+\infty} E_n|F) = \sum_{n=1}^{+\infty} P(E_n|F)$;
3. $k \geq 2$ e $E_1, E_2, ...E_k$ sono eventi di Ω a due a due disgiunti, allora $P(\bigcup_{n=1}^{k} E_n|F) = \sum_{n=1}^{k} P(E_n|F)$;

[2] Per le dimostrazioni delle prime due vedi **S.M. Ross**, *A First Course in Probability, 6th Edition*, Prentice Hall, Upper Saddle River (USA), 2002, pag.96. Le altre tre si dimostrano applicando la definizione di probabilità condizionata.

4. $P(\varnothing|F) = 0$;
5. $P(\overline{E}|F) = 1 - P(E|F)$.

Esempio 2.3.1. Supponiamo di lanciare due dadi, uno per volta. Qual è la probabilità che la somma sia 6, sapendo che è uscito 4 dal primo dado? Sia E l'evento "la somma dei due dadi è 6" e sia F l'evento "Il primo dado è 4". Descriviamo F come l'insieme $F = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$. Quando si lanciano due dadi gli esiti possibili sono 36, cioè Ω ha 36 elementi, e questi eventi hanno tutti probabilità $\frac{1}{36}$. Pertanto $P(F) = \frac{1}{6}$. L'evento EF è "Il primo dado è 4 e la somma dei due dadi è 6", cioè $EF = \{(4,2)\}$. Abbiamo dunque $P(EF) = \frac{1}{36}$. Infine

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{1}{36} : \frac{1}{6} = \frac{1}{6}.$$

Esempio 2.3.2. Supponiamo che un'urna contenga m palline, di cui n sono bianche e le rimanenti m-n sono nere. Eseguiamo due estrazioni successive senza reinserimento. Con quale probabilità si verifica l'evento $E =$ "La seconda pallina estratta è bianca"? Dipende dall'esito della prima estrazione. Sia F l'evento "La prima pallina estratta è bianca" (allora $\overline{F} =$ "La prima pallina estratta è nera").

Abbiamo $P(E|F) = \frac{n-1}{m-1}$ e $P(E|\overline{F}) = \frac{n}{m-1}$. Infatti $P(F) = \frac{n}{m}$ mentre $P(\overline{F}) = \frac{m-n}{m}$. Per le probabilità delle intersezioni abbiamo $P(EF) = \frac{n}{m} \cdot \frac{n-1}{m-1}$ e $P(E\overline{F}) = \frac{m-n}{m} \cdot \frac{n}{m-1}$.

Valgono le seguenti proposizioni:

Proposizione 2.3.1. (Formula della Probabilità Totale) [3] *Dati in uno spazio discreto di probabilità (Ω, P) un evento $E \subseteq \Omega$ e $F_1, F_2, ..., F_n$ una sequenza finita o infinita di eventi a due a due disgiunti, tali che $P(F_i) > 0$ per ogni $i \in \{1, 2, ..., n\}$ e $\bigcup_{i=1}^n F_i = \Omega$, risulta sempre*

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i).$$

[3]Vedi **N. Cufaro Petroni**, *Lezioni di Calcolo delle Probabilità*, Edizioni dal Sud, Modugno(Bari), 1996, pagg.39-40.

Esempio 2.3.3. Consideriamo ancora l'urna dell'esempio 2.3.2, estraiamo in successione e senza reinserimento due palline e, senza guardare la prima, ci chiediamo quale è la probabilità che la seconda sia bianca. Consideriamo ancora gli eventi E e F definiti come sopra. Vogliamo dunque conoscere $P(E)$. La nostra successione di eventi disgiunti si limita pertanto a F e $\overline{F}$ e la Formula della Probabilità Totale diventa: $P(E) = P(E|F)P(F) + P(E|\overline{F})P(\overline{F})$. Abbiamo già calcolato tutto nell'esempio 2.3.2, dunque

$$P(E) = \frac{n-1}{m-1} \cdot \frac{n}{m} + \frac{n}{m-1} \cdot \frac{m-n}{m} = \frac{n}{m}.$$

Mentre $P(E|F)$ e $P(E|\overline{F})$ sono diversi da $P(F)$, risulta $P(E) = P(F)$: quando il primo risultato è sconosciuto esso non influenza la probabilità del secondo.

Proposizione 2.3.2. (Formula di Moltiplicazione)[3] *Comunque assegnati gli eventi $F_1, F_2, ..., F_n$ in uno spazio discreto di probabilità (Ω, P) risulta sempre*

$$P(F_1F_2...F_n) = P(F_n|F_{n-1}...F_1)P(F_{n-1}|F_{n-2}...F_1)...P(F_2|F_1)P(F_1)$$

se $P(F_1F_2...F_{n-1}) > 0$.

Proposizione 2.3.3. (Teorema di Bayes)[3] *Dati due eventi E e F in uno spazio discreto di probabilità (Ω, P) tali che $P(E) > 0, P(F) > 0$, risulta*

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}\ ;$$

inoltre se $F_1, F_2, ..., F_n$ è una sequenza finita o infinita di eventi a due a due disgiunti, tali che $P(F_i) > 0$ per ogni $i \in \{1, 2, ..., n\}$, e per i quali valga che $E \subseteq F_1 \cup F_2 \cup ... \cup F_n$ risulta

$$P(F_i|E) = \frac{P(E|F_i)P(F_i)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}\ .$$

Esempio 2.3.4. Consideriamo una macchina che riconosce come false il 98% delle banconote false e, per errore, l'1% di quelle vere. Si sa che il 2‰ delle banconote da 50 euro in circolazione sono false. Qual è la probabilità che una banconota rilevata come falsa sia in realtà vera?

Sia E l'evento "La banconota è vera" e F l'evento "La banconota è riconosciuta come falsa". Vogliamo conoscere $P(E|F)$ usando la formula di

Bayes. Determiniamo i valori delle probabilità che compaiono nella formula: $P(F|E) = 0.01$, $P(E) = 1 - 0.002 = 0.998$, $P(F) = P(F|E)P(E) + P(F|\overline{E})P(\overline{E}) = 0.01 \cdot 0.998 + 0.98 \cdot 0.002 = 0.01194$. Ora abbiamo tutto:

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)} = \frac{0.01 \cdot 0.998}{0.01194} = 0.836\,.$$

Dato uno spazio di probabilità discreto (Ω, P), diciamo che E e F sono due **eventi indipendenti** se $P(EF) = P(E)P(F)$. Nel caso in cui $P(E) > 0$ e $P(F) > 0$, questo è equivalente a dire che $P(E|F) = P(E)$ e anche $P(F|E) = P(F)$. E e F sono indipendenti se la conoscenza che F si è verificato non influenza la probabilità che si verifichi E, e viceversa. Più in generale, gli eventi $E_1, E_2, ..., E_n$ si dicono indipendenti se per ogni $k = 1, ..., n$ e per ogni insieme di indici $\{i_1, ..., i_k\}$ con $1 \leq i_1 < i_2 < ... < i_k \leq n$ risulta $P(E_{i_1}E_{i_2}...E_{i_k}) = P(E_{i_1})P(E_{i_2})...P(E_{i_k})$, cioè se sono indipendenti due a due, tre a tre,...,n a n.

Esempio 2.3.5. Supponiamo di lanciare due dadi non truccati. Sia E_1 l'evento "La somma dei dadi è sei" e sia F l'evento "Il primo dado risulta quattro". Abbiamo $P(E_1F) = P\{(4,2)\} = \frac{1}{36}$ mentre $P(E_1)P(F) = \frac{5}{36} \cdot \frac{1}{6} = \frac{5}{216}$: infatti $E_1 = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$. Dunque E_1 e F non sono indipendenti. Sia E_2 l'evento "La somma dei dadi è sette".
Abbiamo $P(E_2F) = P\{(4,3)\} = \frac{1}{36}$ e $P(E_2)P(F) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$; infatti $E_2 = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$. Dunque E_2 e F sono indipendenti. In effetti nel primo caso, se il primo dado risultasse sei non avremmo alcuna possibilità di ottenere come somma sei, ma sicuramente un numero più grande di sei. Nel secondo caso invece qualsiasi numero esca, rimane ancora la possibilità col lancio del secondo dado di ottenere sette come somma.

2.4 Variabili aleatorie discrete

Accade che nel fare un esperimento siamo più interessanti a qualche funzione dell'esito rispetto all'esito stesso. Ad esempio nel caso del lancio dei due dadi ci interessa sapere che la somma è sette, ma non in che modo è stata ottenuta, cioè se l'esito dei lanci è stato (1,6) o (2,5) o altro ancora. Queste quantità di interesse, o meglio, queste funzioni a valori reali definite in uno spazio campionario Ω sono dette **variabili aleatorie**. Chiamiamo **variabile aleatoria discreta** una funzione definita in uno spazio campionario Ω e che assume valori in un insieme A finito o numerabile: $X : \Omega \longmapsto A$.

Esempio 2.4.1. Supponiamo di lanciare due dadi non truccati. Definiamo la variabile aleatoria X come la somma dei due dadi. Allora si ha:

- $P\{X = 2\} = P\{(1, 1)\} = \frac{1}{36}$;
- $P\{X = 5\} = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = \frac{4}{36}$;
- $P\{X = 8\} = P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = \frac{5}{36}$;
- $P\{X = 11\} = P\{(5, 6), (6, 5)\} = \frac{2}{36}$.

La variabile X può assumere tutti i valori compresi tra due e dodici e calcolando $P\{X = n\}$ $\forall n \in \{2, 3, ..., 12\}$ otteniamo:

$$P\{\bigcup_{n=2}^{12}\{X = n\}\} = \sum_{n=2}^{12} P\{X = n\} = 1.$$

Esempio 2.4.2. Supponiamo di lanciare due monete non truccate. Definiamo la variabile aleatoria Y come il numero di teste che appaiono. Allora la variabile Y può assumere un valore tra $0, 1, 2$ con le seguenti probabilità (C=croce, T=testa):

- $P\{Y = 0\} = P\{(C, C)\} = \frac{1}{4}$;
- $P\{Y = 1\} = P\{(C, T), (T, C)\} = \frac{2}{4}$;
- $P\{Y = 2\} = P\{(T, T)\} = \frac{1}{4}$.

Calcoliamo subito: $P\{Y = 0\} + P\{Y = 1\} + P\{Y = 2\} = 1$.

Dato un evento $E \subseteq \Omega$ è sempre possibile definire una variabile aleatoria semplice ma importante che prende il nome di **indicatore dell'evento** E $I_E : \Omega \longmapsto \{0, 1\}$, definita nel seguente modo:

$$I_E(\omega) = \begin{cases} 1 & \text{se } \omega \in E \\ 0 & \text{se } \omega \notin E \end{cases}$$

Sia X una variabile discreta a valori in un insieme E. Chiamiamo **densità discreta** della variabile aleatoria X la funzione $p_X : E \longmapsto [0, 1]$, e $p_X(x) = P\{X = x\}$. Chiamiamo **distribuzione o legge** della variabile aleatoria X la funzione $\mu_X : \mathcal{P}(E) \longmapsto [0, 1]$, e $\mu_X(A) = P\{X \in A\}$. Per la funzione μ_X valgono le seguenti proprietà:

1. se A_n è una successione di sottoinsiemi a due a due disgiunti di E, allora $\mu_X(\bigcup\limits_n A_n) = \sum\limits_n \mu_X(A_n)$;

2. per ogni sottoinsieme A di E si ha $\mu_X(A) = \sum\limits_{x \in A} P\{X = x\} = \sum\limits_{x \in A} \mu_X(\{x\})$.

Tra la funzione p_X e μ_X sussistono le seguenti relazioni: $p_X(x) = \mu_X(\{x\})$ e $\mu_X(A) = \sum_{x \in A} p_X(x)$. Se X è una variabile vettoriale (o vettore aleatorio), assume cioè valori in $\mathbb{R}^n$, allora $X = (X_1, X_2, ..., X_n)$ e chiamiamo $p_X(x) = p_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$ **densità congiunta** delle variabili $X_1, X_2, ..., X_n$. Chiamiamo invece **densità marginali** le densità delle componenti, cioè $p_{X_1}(x_1), p_{X_2}(x_2), ..., p_{X_n}(x_n)$.
Vale la seguente proprietà:

- se $k \in \{1, 2, ..., n\}$, allora

$$p_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = \sum_{x_{k+1},...,x_n \in \mathbb{R}} p_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) \ .$$

È possibile combinare delle variabili aleatorie per ottenere nuove variabili aleatorie introducendo il concetto di funzione di una variabile aleatoria o più in generale funzione di un vettore aleatorio. Dato un vettore aleatorio X con n componenti e data una funzione reale di n variabili $f : \mathbb{R}^n \longmapsto \mathbb{R}$, è possibile definire una nuova variabile aleatoria Y mediante la posizione $Y = f[X_1, X_2, ..., X_n]$.

Sia $X = (X_1, X_2, ..., X_n)$ una variabile aleatoria vettoriale che assume valori in $E = E_1 \times E_2 \times ... \times E_n$. Le componenti di X si dicono **indipendenti** se per ogni scelta di $B_1 \subseteq E_1, B_2 \subseteq E_2, ..., B_n \subseteq E_n$ si verifica:

$$P\{X_1 \in B_1, X_2 \in B_2, ..., X_n \in B_n\} = \prod_{i=1}^{n} P\{X_i \in B_i\} \ .$$

Il concetto di indipendenza delle componenti di una variabile vettoriale di dimensione n si traduce immediatamente nel concetto di indipendenza per n variabili aleatorie. Si può dare anche una definizione più rigorosa di **variabili aleatorie indipendenti**: sia I un insieme qualunque di indici e $\{X_i : i \in I\}$ una famiglia di variabili aleatorie a valori negli insiemi $E_i, i \in I$. Le variabili aleatorie di questa famiglia si dicono indipendenti se per ogni $J \subset I$ finito e per ogni scelta di $B_j \subseteq E_j, j \in J$, si ha

$$P(\bigcap_{j \in J} \{X_j \in B_j\}) = \prod_{j \in J} P\{X_j \in B_j\} \ .$$

Proposizione 2.4.1. *Valgono i seguenti risultati:*

1. *se le variabili* $X_i : i \in I$ *sono indipendenti allora per ogni scelta di* $B_i \subseteq E_i, i \in I$*, gli eventi* $\{X_i \in B_i\}, i \in I$*, sono indipendenti e viceversa;*

2. *se* $X_1, X_2, ..., X_n$ *sono variabili aleatorie definite in uno stesso spazio di probabilità* (Ω, P)*, a valori rispettivamente negli insiemi* $E_1, E_2, ..., E_n$*, allora* $X_1, X_2, ..., X_n$ *sono indipendenti se e solo se*

$$p_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p_{X_i}(x_i);$$

3. *siano* $X_1, X_2, ..., X_n$ *variabili aleatorie indipendenti e* $I = \{i_1, i_2, ..., i_h\}$ *e* $J = \{j_1, j_2, ..., j_k\}$ *due sottoinsiemi non vuoti e disgiunti di* $\{1, 2, ..., n\}$*. Indichiamo con* X_I *e* X_J *le seguenti variabili aleatorie vettoriali:* $X_I = (X_{i_1}, X_{i_2}, ..., X_{i_h})$ *e* $X_J = (X_{j_1}, X_{j_2}, ..., X_{j_k})$*.* X_I *e* X_J *sono variabili aleatorie indipendenti;*

4. *siano* X *e* Y *due variabili aleatorie definite in* (Ω, P)*, a valori rispettivamente negli insiemi* E *e* F*; siano* H *e* K *due insiemi e* $f : E \longmapsto H, g : F \longmapsto K$ *funzioni arbitrarie. Se* X *e* Y *sono indipendenti allora anche* $f(X)$ *e* $g(Y)$ *sono indipendenti.*

Assegnata su (Ω, P) una variabile aleatoria discreta X, si chiama **funzione di distribuzione o di ripartizione** di X la seguente funzione, definita $\forall x \in \mathbb{R}$:

$$F_X(x) = P\{\omega \in \Omega : X(\omega) \leq x\} = P\{X \leq x\} = \sum_{i \leq x} p_X(i).$$

Se $Y = (Y_1, Y_2, ..., Y_n)$ è un vettore aleatorio, chiamiamo **funzione di distribuzione congiunta** di Y la funzione

$$F_Y(y_1, y_2, ..., y_n) = P\{\omega \in \Omega : Y_1(\omega) \leq y_1, ..., Y_n(\omega) \leq y_n\}.$$

Alcune proprietà di $F_X(x)$ sono [4]:

1. $F_X(x)$ è funzione non decrescente di x ;

2. $\lim\limits_{x \to +\infty} F_X(x) = 1$;

[4]Per le prime quattro vedi **S.M. Ross**, *A First Course in Probability, 6th Edition*, Prentice Hall, Upper Saddle River (USA), 2002, pag.167. Per la quinta vedi **G.R. Grimmett, D. Stirzaker**, *Probability and Random Processes*, Clarendon Press, Oxford, 1982, pag.21. Per la sesta basta applicare la definizione di componenti indipendenti di un vettore aleatorio.

3. $\lim_{x \to -\infty} F_X(x) = 0$;
4. è continua da destra: se x_n è una successione decrescente che tende a x allora $\lim_{n \to +\infty} F_X(x_n) = F_X(x)$;
5. per ogni $x \in \mathbb{R}$ vale la relazione $p_X(x) = F_X(x) - F_X(x^-)$, dove $F_X(x^-) = \lim_{y \to x^-} F_X(y)$;
6. se le componenti di $Y = (Y_1, ..., Y_n)$ sono indipendenti allora $F_Y(y_1, y_2, ..., y_n) = F_{Y_1}(y_1) \cdots F_{Y_n}(y_n)$.

2.4.1 Esempi di variabili aleatorie discrete

La variabile aleatoria di Bernoulli Supponiamo di fare una prova o un esperimento il cui esito possa essere classificato come "successo" o "insuccesso". Sia X variabile aleatoria definita nel seguente modo:

$$X = \begin{cases} 1 & \text{se l'esito è un successo} \\ 0 & \text{se l'esito è un insuccesso} \end{cases}$$

Sia $p \in]0,1[$ la probabilità che la prova sia "un successo". Descriviamo la funzione p_X nel seguente modo:

$$p_X(0) = P\{X = 0\} = 1 - p$$

$$p_X(1) = P\{X = 1\} = p.$$

Per indicare che X è una variabile aleatoria di Bernoulli di parametro p, scriviamo $X \sim Be(p)$.

La variabile aleatoria binomiale Supponiamo di fare n prove indipendenti e supponiamo che l'esito di ciascuna prova sia classificabile come "successo" o "insuccesso" e che la probabilità di successo sia $p \in]0,1[$. Se X rappresenta il numero di successi verificati in n prove, allora X è detta variabile aleatoria binomiale di parametri (n,p) e si scrive $X \sim B(n,p)$. La densità di X è:

$$p_X(i) = \binom{n}{i} p^i (1-p)^{n-i}.$$

La variabile aleatoria binomiale X ha la stessa distribuzione di una variabile aleatoria del tipo $Y = X_1 + X_2 + ... + X_n$, dove le X_i sono variabili aleatorie di Bernoulli tutte dello stesso parametro p e tra loro indipendenti.

La variabile aleatoria geometrica Supponiamo di fare prove indipendenti, ciascuna con probabilità $p \in]0,1[$ di avere successo, fino a quando si verifica un successo. Se X rappresenta il numero di prove fino al primo successo, allora X è detta variabile aleatoria geometrica di parametro p e si scrive $X \sim Ge(p)$. La densità di X è:

$$p_X(n) = (1-p)^{n-1}p.$$

Si intende che gli insuccessi sono stati $n-1$ e che alla $n-esima$ prova si sia ottenuto il primo successo.

La variabile aleatoria di Poisson Una variabile aleatoria X che assume valore $n = 0, 1, 2, ...$ è detta di Poisson di parametro $\lambda (> 0)$ (e si scrive $X \sim Po(\lambda)$) se la sua densità può essere scritta come:

$$p_X(n) = e^{-\lambda}\frac{\lambda^n}{n!}.$$

Una proprietà importante di questa variabile è che può approssimare una variabile aleatoria binomiale di parametri (n,p), quando n è molto grande e p molto piccolo. Basta porre allora $\lambda = np$ e scrivere la densità di una variabile binomiale Y di parametri $(n, \frac{\lambda}{n})$.

2.5 Valore atteso, momenti, varianza e covarianza

Sia X una variabile aleatoria discreta su uno spazio discreto (Ω, P), e sia p_X la sua densità. Definiamo **valore medio o valore atteso di** X il numero $E[X] = \sum_{x:p_X(x)>0} x p_X(x)$, quando $\sum_{x:p_X(x)>0} |x| p_X(x) < +\infty$. Si tratta di una media ponderata dei possibili valori che X può assumere.

Esempio 2.5.1. Sia X variabile aleatoria definita come l'esito del lancio di un dado non truccato. Vogliamo conoscere $E[X]$. X può assumere un valore compreso tra uno e sei, ciascuno con probabilità $\frac{1}{6}$: $p_X(i) = \frac{1}{6}$, $\forall i \in [1,6]$. Calcoliamo il valore atteso:

$$E[X] = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{7}{2}\,.$$

A volte si è interessati a determinare il valore atteso di una qualche funzione di X, $f(X)$. Essendo $f(X)$ una variabile aleatoria, avrà una sua densità calcolabile a partire dalla densità di X. Una volta ottenuta la densità di $f(X)$, possiamo determinare $E[f(X)]$ usando la definizione di valore atteso.

Esempio 2.5.2. Supponiamo che X abbia la seguente densità:
$p(0) = 0.2$, $p(1) = 0.5$, $p(2) = 0.3$
Vogliamo conoscere $E[f(X)]$,con $f(X) = X^2$. Poniamo $Y = X^2$. Y può assumere un valore tra $0^2, 1^2, 2^2$ con le rispettive probabilità:
$p_Y(0^2) = P\{Y = 0\} = 0.2$
$p_Y(1^2) = P\{Y = 1\} = 0.5$
$p_Y(2^2) = P\{Y = 4\} = 0.3$
Calcoliamo infine $E[Y] = 0(0.2) + 1(0.5) + 4(0.3) = 1.7$.

Esiste una via più semplice che ci permette di determinare $E[f(X)]$ senza prima calcolare la densità di $f(X)$. Questa via è descritta dalla seguente proposizione [5]:

Proposizione 2.5.1. *Se X è una variabile aleatoria discreta con densità p_X, allora per ogni funzione a valori reali f, tale che la variabile $f(X)$ ammetta valore atteso finito, si ha:* $E[f(X)] = \sum\limits_{x:p_X(x)>0} f(x)p_X(x)$.

Esempio 2.5.3. Riprendendo l'esempio precedente, calcoliamo subito $E[X^2] = 0^2(0.2) + 1^2(0.5) + 2^2(0.3) = 1.7$.

Descriviamo le proprietà del valore atteso nella seguente proposizione [6]:

Proposizione 2.5.2. *Siano X e Y due variabili aleatorie definite in (Ω, P) che ammettono valore atteso finito:*

1. *dati due numeri a e b risulta*

$$E[aX + bY] = aE[X] + bE[Y],$$

cioè $E[\cdot]$ è funzione lineare;

2. *se X e Y sono tali che $X(\omega) \leq Y(\omega), \forall\omega \in \Omega$, allora*

$$E[X] \leq E[Y];$$

3. *per ogni variabile aleatoria X è verificata la seguente relazione:*

$$|E[X]| \leq E[|X|];$$

[5] Vedi **S.M. Ross**, *A First Course in Probability, 6th Edition*, Prentice Hall, Upper Saddle River (USA), 2002, pag.134.

[6] Per le prime cinque vedi **N. Cufaro Petroni**, *Lezioni di Calcolo delle Probabilità*, Edizioni dal Sud, Modugno(Bari), 1996, pagg.63-65, in particolare la 5 è la generalizzazione del caso n=2, la 6 è ovvia.

4. *se A è un evento di Ω e I_A il suo indicatore, allora $E[I_A] = P(A)$;*

5. *se $X_1, X_2, ..., X_n$ sono variabili aleatorie indipendenti definite in (Ω, P) allora è verificata la seguente relazione:*

$$E[X_1 \cdot X_2 \cdots X_n] = E[X_1] \cdot E[X_2] \cdots E[X_n];$$

6. *se a è una costante allora $E[a] = a$.*

Ci si riferisce al valore atteso della variabile aleatoria X anche con il nome di **momento primo** di X. La quantità $E[X^n], n \geq 1$, quando esiste finita, si chiama **momento n-simo o momento di ordine n** di X. Per la proposizione 2.5.1, siamo in grado di calcolare: $E[X^n] = \sum\limits_{x:p_X(x)>0} x^n p_X(x)$.
Un'altra quantità di interesse è la **varianza** di una variabile aleatoria X, indicata con $Var(X)$ e definita da $Var(X) = E[(X - E[X])^2]$. La varianza di X misura il valore atteso del quadrato della deviazione di X dal suo valore atteso. Osserviamo che:

$$Var(X) = E[(X - E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2] =$$

$$= E[X^2] - 2E[X]E[X] + (E[X])^2 = E[X^2] - (E[X])^2.$$

Esempio 2.5.4. Calcoliamo $Var(X)$ quando X rappresenta l'esito del lanci di un dado non truccato. Usiamo l'ultima relazione: abbiamo bisogno di $E[X]$ e di $E[X^2]$. Il primo è stato calcolato nell'esempio 2.5.1: $E[X] = \frac{7}{2}$. Per l'altro: $E[X^2] = 1(\frac{1}{6}) + 4(\frac{1}{6}) + 9(\frac{1}{6}) + 16(\frac{1}{6}) + 25(\frac{1}{6}) + 36(\frac{1}{6}) = (\frac{91}{6})$. Possiamo scrivere infine: $Var(X) = (\frac{91}{6}) - (\frac{7}{2})^2 = \frac{35}{12}$.

La varianza non è un operatore lineare. Si dimostra infatti che [7]:

Proposizione 2.5.3. *Siano X e Y due variabili aleatorie definite in (Ω, P) che ammettono momento secondo e siano a e b due numeri arbitrari:*

1. $Var(a) = 0$;

2. $Var(bX) = b^2 Var(X)$;

3. $Var(a + bX) = b^2 Var(X)$;

4. *se anche Y è una variabile aleatoria, e X, Y sono indipendenti allora $Var(X + Y) = Var(X) + Var(Y)$ (vale anche per n variabili indipendenti).*

[7] Per le prime tre vedi **N. Cufaro Petroni**, *Lezioni di Calcolo delle Probabilità*, Edizioni dal Sud, Modugno(Bari), 1996, pagg.68-69. Per l'ultima vedi **G.R. Grimmett, D. Stirzaker**, *Probability and Random Processes*, Clarendon Press, Oxford, 1982, pag.34.

Per conoscere $Var(X+Y)$ quando X e Y non sono indipendenti occorre introdurre il concetto di covarianza. Date due variabili aleatorie X e Y chiamiamo **covarianza** il numero $Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$. Come abbiamo fatto con la varianza cerchiamo un altro modo per esprimere la covarianza:

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])] =$$

$$= E[XY - XE[Y] - YE[X] + E[X]E[Y]] = E[XY] - E[X]E[Y] +$$
$$-E[Y]E[X] + E[X]E[Y] = E[XY] - E[X]E[Y] \ .$$

Da questa relazione notiamo subito che se X e Y sono variabili aleatorie indipendenti allora $Cov(X,Y) = 0$. Quando $Cov(X,Y) = 0$ si dice che le variabili sono **non correlate**. È vero che se due variabili sono indipendenti allora sono non correlate, ma non il viceversa. Vediamo le proprietà della covarianza:

Proposizione 2.5.4. *Siano X, Y e Z variabili aleatorie, definite in (Ω, P), che ammettono momento d'ordine due finito e sia c una costante. Valgono* [8]*:*

1. $Cov(X,X) = Var(X)$;
2. $Cov(X,Y) = Cov(Y,X)$;
3. $Cov(cX,Y) = cCov(X,Y)$;
4. $Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$;
5. *in generale:* $Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.

Possiamo calcolare $Var(X+Y) = E\{[(X+Y) - E[X+Y]]^2\} = E\{[(X - E[X]) + (Y-E[Y])]^2\} = E\{(X-E[X])^2 + (Y-E[Y])^2 + 2(X-E[X])(Y-E[Y])\} = Var(X) + Var(Y) + 2Cov(X,Y)$.

2.5.1 Calcoli di valore atteso e varianza

Variabile aleatoria di Bernoulli Sia X di parametro p, calcoliamo il valore atteso: $E[X] = 0 \cdot (1-p) + 1 \cdot p = p$. Per calcolare la varianza, dobbiamo conoscere anche il momento secondo: $E[X^2] = 0^2 \cdot (1-p) + 1^2 \cdot p = p$. Infine ecco la varianza: $Var(X) = p - p^2$.

[8] Le prime tre sono immediate, per la quarta si veda **S.M. Ross**, *Introduction to Probability Models, Eighth Edition*, Academic Press, USA, 2003, pag.54. La quinta è conseguenza immediata della quarta.

Variabile aleatoria binomiale Sia X di parametri (n, p), per quanto detto nel paragrafo 2.4.1, si ha $E[X] = E[X_1 + X_2 + ... + X_n] = nE[X_1] = np$ e $Var(X) = Var(X_1 + X_2 + ... + X_n) = nVar(X_1) = n(p - p^2)$.

Variabile aleatoria geometrica Sia X di parametro p. $E[X] = \sum_{n=1}^{+\infty} np(1-p)^{n-1} = \frac{p}{1-p}\sum_{n=1}^{+\infty} n(1-p)^n = \frac{p}{1-p} \cdot \frac{1-p}{p^2} = \frac{1}{p}$. Siamo arrivati a questo risultato sfruttando la somma della serie geometrica. Posto $q = 1 - p \in]0, 1[$, è noto che $\sum_{n=0}^{+\infty} q^n = \frac{1}{1-q}$ e derivando l'uguaglianza si ottiene: $\sum_{n=1}^{+\infty} nq^{n-1} = \frac{1}{(1-q)^2}$ e infine $\sum_{n=1}^{+\infty} nq^n = \frac{q}{(1-q)^2}$. Derivando ancora una volta quest'ultima otteniamo: $\sum_{n=1}^{+\infty} n^2q^n = \frac{q(1+q)}{(1-q)^3}$. Calcoliamo il momento secondo di X: $E[X^2] = \sum_{n=1}^{+\infty} n^2p(1-p)^{n-1} = \frac{p}{1-p}\sum_{n=1}^{+\infty} n^2(1-p)^n = \frac{p}{1-p}\frac{(1-p)(2-p)}{p^3} = \frac{2-p}{p^2}$. La varianza risulta dunque: $Var(X) = \frac{1-p}{p^2}$.

Variabile aleatoria di Poisson Sia X di parametro λ. $E[X] = \sum_{n=1}^{+\infty} ne^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=1}^{+\infty}\frac{\lambda^n}{(n-1)!} = \lambda e^{-\lambda}\sum_{n=1}^{+\infty}\frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$. Per il momento secondo troviamo $E[X^2] = \sum_{n=1}^{+\infty} n^2e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=1}^{+\infty}[n(n-1)+n]\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=2}^{+\infty}\frac{\lambda^n}{(n-2)!} + E[X] = e^{-\lambda}\lambda^2\sum_{n=2}^{+\infty}\frac{\lambda^{n-2}}{(n-2)!} + \lambda = \lambda^2 + \lambda$. Infine la varianza: $Var(X) = \lambda$.

Capitolo 3

Valore atteso condizionato: caso discreto

Nel secondo capitolo abbiamo parlato della probabilità condizionata:
dato uno spazio di probabilità discreto (Ω, P) e due eventi $E, F \subseteq \Omega$ tali che $P(F) > 0$, chiamiamo **probabilità condizionata** di E rispetto a F, ossia la probabilità che si verifichi E sapendo che si è verificato F, la seguente quantità:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(EF)}{P(F)} \quad .$$

Abbiamo parlato anche di valore atteso di una variabile aleatoria discreta:
se X è una variabile aleatoria discreta su uno spazio discreto (Ω, P), e se p_X è la sua densità, chiamiamo **valore medio o valore atteso di** X il numero $E[X] = \sum_{x:p_X(x)>0} xp_X(x)$, nell'ipotesi in cui $\sum_{x:p_X(x)>0} |x|p_X(x) < +\infty$. Vogliamo ora determinare il valore atteso condizionato.

3.1 Densità condizionata

Siano X e Y variabili aleatorie discrete definite in uno stesso spazio (Ω, P). Chiamiamo densità condizionata di X dato $Y = y$, (per tutti i valori di y tali che $P\{Y = y\} > 0$):

$$p_{X|Y}(x|y) = P\{X = x|Y = y\} = \frac{P\{X = x, Y = y\}}{P\{Y = y\}} = \frac{p_{X,Y}(x, y)}{p_Y(y)} \quad .$$

Se X e Y sono indipendenti risulta:

$$p_{X|Y}(x|y) = P\{X = x|Y = y\} = P\{X = x\}.$$

Allo stesso modo definiamo la funzione di ripartizione condizionata di X dato $Y=y$, (per tutti i valori di y tali che $P\{Y=y\}>0$), come:

$$F_{X|Y}(x|y)=P\{X\leq x|Y=y\}=\sum_{a\leq x}p_{X|Y}(a|y)\ .$$

Esempio 3.1.1. Supponiamo che la densità congiunta delle variabili X e Y sia data da:
$p_{X,Y}(1,1)=0.5\quad p_{X,Y}(1,2)=0.1\ \ p_{X,Y}(2,1)=0.1\ \ p_{X,Y}(2,2)=0.3.$
Calcoliamo la densità condizionata di X dato $Y=1$. Abbiamo che:

$$p_Y(1)=\sum_x p_{X,Y}(x,1)=p_{X,Y}(1,1)+p_{X,Y}(2,1)=0.5+0.1=0.6$$

e quindi

$$p_{X|Y}(1|1)=P\{X=1|Y=1\}=\frac{P\{X=1,Y=1\}}{P\{Y=1\}}=\frac{p_{X,Y}(1,1)}{p_Y(1)}=\frac{0.5}{0.6}=\frac{5}{6}$$

$$p_{X|Y}(2|1)=P\{X=2|Y=1\}=\frac{P\{X=2,Y=1\}}{P\{Y=1\}}=\frac{p_{X,Y}(2,1)}{p_Y(1)}=\frac{0.1}{0.6}=\frac{1}{6}\ .$$

Esempio 3.1.2. Siano X e Y variabili aleatorie di Poisson indipendenti, rispettivamente di parametri λ e μ. Calcoliamo la probabilità condizionata di X, dato $X+Y=n$.

$$P\{X=k|X+Y=n\}=\frac{P\{X=k,X+Y=n\}}{P\{X+Y=n\}}=\frac{P\{X=k,Y=n-k\}}{P\{X+Y=n\}}$$

Determiniamo separatamente il denominatore. Osserviamo che $X+Y$ è una variabile di Poisson di parametro $\lambda+\mu$:
$P\{X+Y=n\}=\sum_{k=0}^{n}P\{X=k,Y=n-k\}=\sum_{k=0}^{n}P\{X=k\}P\{Y=n-k\}=$
$\sum_{k=0}^{n}e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}=e^{-(\lambda+\mu)}\sum_{k=0}^{n}\frac{\lambda^k\mu^{n-k}}{k!(n-k)!}=\frac{e^{-(\lambda+\mu)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda^k\mu^{n-k}=$
$=\frac{e^{-(\lambda+\mu)}}{n!}(\lambda+\mu)^n.$
Calcoliamo infine la probabilità richiesta:

$$P\{X=k|X+Y=n\}=\frac{P\{X=k,Y=n-k\}}{P\{X+Y=n\}}=$$

$$=e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}\frac{n!}{e^{-(\lambda+\mu)}(\lambda+\mu)^n}=\binom{n}{k}\frac{\lambda^k\mu^{n-k}}{(\lambda+\mu)^n}=$$

$$= \binom{n}{k} (\frac{\lambda}{\lambda+\mu})^k (\frac{\mu}{\lambda+\mu})^{n-k} = \binom{n}{k} (\frac{\lambda}{\lambda+\mu})^k (1-\frac{\lambda}{\lambda+\mu})^{n-k}.$$

Questa è una distribuzione binomiale di parametri n e $\lambda/(\lambda+\mu)$.

3.2 Valore atteso condizionato

Il **valore atteso condizionato** della variabile X dato $Y = y$ è definito da:

$$E[X|Y=y] = \sum_x xP\{X=x|Y=y\} = \sum_x xp_{X|Y}(x|y).$$

Come ipotesi chiederemo sempre che X ammetta valore atteso finito.
Le sue proprietà sono descritte nella proposizione 5.2.3 nel caso generale.

Esempio 3.2.1. Consideriamo un esperimento che abbia tre esiti possibili indicati con 1,2,3, i quali hanno rispettivamente probabilità di successo p_1, p_2, p_3 tali che $p_1 + p_2 + p_3 = 1$. Supponiamo di fare n prove indipendenti di questo esperimento. Indichiamo con X_i, $i = 1, 2, 3$ il numero di volte che si verifica l'esito i. Vogliamo calcolare il valore atteso di X_1 dato $X_2 = m$.
Innanzitutto determiniamo la densità. Per $k \leq n - m$,

$$P\{X_1 = k|X_2 = m\} = \frac{P\{X_1 = k, X_2 = m\}}{P\{X_2 = m\}}.$$

Se $X_1 = k$ e $X_2 = m$ allora $X_3 = n - m - k$.

$$P\{X_1 = k, X_2 = m, X_3 = n-k-m\} = \frac{n!}{k!m!(n-k-m)!} p_1^k p_2^m p_3^{n-k-m}.$$

Infatti vengono eseguiti n esperimenti nei quali l'esito 1 appare k volte, l'esito 2, m volte e l'esito 3 le rimanenti $n-k-m$ volte. Pertanto una qualsiasi sequenza di questo tipo ha probabilità di verificarsi $p_1^k p_2^m p_3^{n-k-m}$. Quante sono queste sequenze? Immaginiamo di avere n scatolette disposte in fila, uno accanto all'altra. Supponiamo di avere n caramelle di cui k alla fragola, m alla menta e le rimanenti alla liquirizia. Vogliamo mettere in ogni scatoletta una sola caramella. Posizioniamo prima quelle alla fragola. In quanti modi lo possiamo fare? Si tratta di una combinazione di n elementi presi k per volta, e quindi i modi possibili sono $\binom{n}{k}$. Ora posizioniamo quelle alla menta: ci rimangono a disposizione $n - k$ scatolette. Possiamo quindi sistemare queste caramelle in

$\binom{n-k}{m}$ modi. Infine mettiamo nelle scatolette le caramelle alla liquirizia, e qui non abbiamo più scelta. Rispondiamo ora alla domanda iniziale: possiamo posizionare le n caramelle in $\binom{n}{k}\binom{n-k}{m}$ modi. Quindi le sequenze possibili sono

$$\binom{n}{k}\binom{n-k}{m} = \frac{n!}{k!(n-k)!}\frac{(n-k)!}{m!(n-k-m)!} = \frac{n!}{k!m!(n-k-m)!}.$$

Si dice che il vettore aleatorio (X_1, X_2, X_3) ha legge multinomiale.
Proprio perché il terzo esito è determinato in conseguenza agli esiti 1 e 2, possiamo scrivere $P\{X_1 = k, X_2 = m\} = P\{X_1 = k, X_2 = m, X_3 = n-k-m\}$. Osserviamo anche che X_2 è una variabile binomiale di parametri n e p_2.
Calcoliamo infine la probabilità condizionata:

$$P\{X_1 = k|X_2 = m\} = \frac{\frac{n!}{k!m!(n-k-m)!}p_1^k p_2^m p_3^{n-k-m}}{\frac{n!}{m!(n-m)!}p_2^m(1-p_2)^{n-m}} =$$

$$= \frac{(n-m)!}{k!(n-k-m)!}(\frac{p_1}{1-p_2})^k(\frac{p_3}{1-p_2})^{n-k-m} =$$

$$= \binom{n-m}{k}(\frac{p_1}{1-p_2})^k(1-\frac{p_1}{1-p_2})^{n-k-m}.$$

Questa distribuzione è una binomiale di parametri $n-m$ e $p_1/(1-p_2)$. Sappiamo che il valore atteso di una variabile binomiale di parametri n e p è np. In questo caso abbiamo dunque:

$$E[X_1|X_2 = m] = (n-m)\frac{p_1}{1-p_2}$$

Esempio 3.2.2. Una scatola contiene tre palline bianche, sei rosse e cinque nere. Si estraggono casualmente dalla scatola sei palline, e ciascuna dopo essere stata estratta viene rimessa nella scatola. Denotiamo con X il numero di palline bianche estratte, e con Y il numero di palline nere estratte. Qual è il valore atteso di X condizionato a $Y = 3$?
Determiniamo innanzitutto $p_{X|Y}(k|3) = P\{X = k|Y = 3\} = \frac{P\{X=k,Y=3\}}{P\{Y=3\}}$. X può assumere i valori 0,1,2,3. Calcoliamo separatamente i vari casi:

$$p_{X|Y}(0|3) = \frac{\binom{6}{3}(\frac{5}{14})^3(\frac{6}{14})^3}{\binom{6}{3}(\frac{5}{14})^3(\frac{9}{14})^3} = \frac{8}{27}$$

$$p_{X|Y}(1|3) = \frac{\binom{6}{1}\binom{5}{3}\frac{3}{14}(\frac{5}{14})^3(\frac{6}{14})^2}{\binom{6}{3}(\frac{5}{14})^3(\frac{9}{14})^3} = \frac{4}{9}$$

$$p_{X|Y}(2|3) = \frac{\binom{6}{2}\binom{4}{3}(\frac{3}{14})^2(\frac{5}{14})^3\frac{6}{14}}{\binom{6}{3}(\frac{5}{14})^3(\frac{9}{14})^3} = \frac{2}{9}$$

$$p_{X|Y}(3|3) = \frac{\binom{6}{3}(\frac{3}{14})^3(\frac{5}{14})^3}{\binom{6}{3}(\frac{5}{14})^3(\frac{9}{14})^3} = \frac{1}{27}\,.$$

Calcoliamo il valore atteso condizionato:

$$E[X|Y=3] = 0\cdot p_{X|Y}(0|3) + 1\cdot p_{X|Y}(1|3) + 2\cdot p_{X|Y}(2|3) + 3\cdot p_{X|Y}(3|3) =$$
$$= \frac{4}{9} + \frac{4}{9} + \frac{1}{9} = 1.$$

Dimostriamo ora un'importante proprietà del valore atteso condizionato:

Proposizione 3.2.1. *Per tutte le variabili aleatorie discrete X e Y, con X avente valore atteso finito, si ha*

$$E[X] = E[E[X|Y]] = \sum_y E[X|Y=y]P\{Y=y\}.$$

Infatti:

$$\sum_y E[X|Y=y]P\{Y=y\} = \sum_y\sum_x xP\{X=x|Y=y\}P\{Y=y\} =$$
$$= \sum_y\sum_x x\frac{P\{X=x,Y=y\}}{P\{Y=y\}}P\{Y=y\} = \sum_y\sum_x xP\{X=x,Y=y\} =$$
$$= \sum_x\sum_y P\{X=x,Y=y\} = \sum_x P\{X=x\} = E[X].$$

Scriviamo $E[E[X|Y]] = \sum_y E[X|Y=y]P\{Y=y\}$ perché devo pensare $E[X|Y]$ come una funzione di Y e cioè del tipo $g(Y)$. Ma noi sappiamo che $E[g(Y)] = \sum_y g(y)p_Y(y)$ e dunque segue quanto scritto sopra.

Questa formula è usata soprattutto per calcolare valori attesi di variabili aleatorie quando risulta più facile calcolare il valore atteso condizionato rispetto a qualche altra variabile. Vediamo subito delle applicazioni.

Esempio 3.2.3. In un museo il valore atteso di visite guidate per settimana è venti. Supponiamo che il numero di visitatori ad ogni visita guidata sia rappresentata da una variabile aleatoria che ha valore atteso dieci. Assumiamo inoltre che il numero di visitatori ad ogni visita sia indipendente dal numero delle visite che ci sono in una settimana. Qual è il numero atteso di visitatori in una settimana?
Indichiamo con N il numero di visite in una settimana e con X_i il numero di visitatori alla visita i-esima. Allora il totale dei visitatori in una settimana è $\sum_{i=1}^{N} X_i$ e il numero atteso che vogliamo conoscere è $E[\sum_{i=1}^{N} X_i]$. Lo calcoliamo usando la proprietà del valore medio condizionato:

$$E[\sum_{i=1}^{N} X_i] = E[E[\sum_{i=1}^{N} X_i|N]] \ .$$

Studiamo la variabile $E[\sum_{i=1}^{N} X_i|N]$:

$$E[\sum_{i=1}^{N} X_i|N=n] = E[\sum_{i=1}^{n} X_i|N=n] = E[\sum_{i=1}^{n} X_i] = nE[X]$$

perché per ipotesi le variabili X_i e N sono indipendenti e X è una qualunque delle X_i. Dunque:

$$E[\sum_{i=1}^{N} X_i|N] = NE[X].$$

Infine:

$$E[E[\sum_{i=1}^{N} X_i|N]] = E[NE[X]] = E[N]E[X] = 20 \cdot 10 = 200.$$

In una settimana ci si aspettano duecento visitatori.

Esempio 3.2.4. In un luna park un bambino sta giocando in un labirinto e si trova di fronte a tre porte: la numero uno porta all'uscita dopo cinque minuti di passeggiata; la numero due porta il bambino di nuovo a quel punto dopo tre minuti di passeggiata; la numero tre porta il bambino di nuovo a quel punto dopo otto minuti di passeggiata. Il bambino non conosce qual è la porta giusta

per uscire, ed è molto distratto, per cui se sbaglia porta e torna di nuovo lì non ricorda la porta già provata. Assumiamo dunque che ogni volta il bambino scelga in modo uguale la porta. Qual è il tempo che in media il bambino impiegherà per uscire dal labirinto?
Denotiamo con X il tempo impiegato per uscire e con Y la porta che il bambino sceglie. Pertanto Y assume i valori 1,2,3. Calcoliamo $E[X]$ mediante la proprietà del valore atteso condizionato:

$$E[X] = E[X|Y = 1]P\{Y = 1\}+E[X|Y = 2]P\{Y = 2\}+E[X|Y = 3]P\{Y = 3\}$$

Abbiamo: $P\{Y = 1\} = P\{Y = 2\} = P\{Y = 3\} = \frac{1}{3}$ e calcoliamo:

$$E[X|Y = 1] = 5$$

$$E[X|Y = 2] = E[3 + X] = 3 + E[X]$$
$$E[X|Y = 3] = E[8 + X] = 8 + E[X].$$

Infatti quando sceglie la seconda o la terza porta cammina rispettivamente per tre e per otto minuti, ma poi si trova di nuovo allo stesso punto per cui aggiungo ancora il tempo per uscire, X. Dunque $X|Y = 2$ è distribuita come $X+3$, $X|Y = 3$ come $X+8$. Scriviamo l'equazione che ha per incognita $E[X]$:

$$E[X] = \frac{1}{3}(5 + 3 + E[X] + 8 + E[X])$$

e quindi $E[X] = 16$.
Il bambino impiegherà in media sedici minuti per uscire dal labirinto.

Esempio 3.2.5. Giochiamo con una moneta e supponiamo che la probabilità che esca testa sia p; continuiamo a lanciare la moneta finché non compare la prima testa. Qual è il numero di lanci medio richiesto?
Denotiamo con N il numero di lanci richiesti e con Y la variabile che indica se esce testa o croce al lancio della moneta. La descriviamo così:

$$Y = \begin{cases} 1 & \text{se al primo lancio esce testa} \\ 0 & \text{se al primo lancio esce croce} \end{cases}$$

Usiamo la solita proprietà per calcolare $E[N]$:

$$E[N] = E[N|Y = 1]P\{Y = 1\} + E[N|Y = 0]P\{Y = 0\}.$$

Per ipotesi $P\{Y = 1\} = p$ e $P\{Y = 0\} = 1 - p$.
Per i valori attesi condizionati abbiamo:

$$E[N|Y = 1] = 1 \qquad E[N|Y = 0] = E[1 + N] = 1 + E[N].$$

Per la prima, quando $Y = 1$ sappiamo che al primo lancio è risultato testa e naturalmente il valore atteso è uno. Per quanto riguarda la seconda invece, se $Y = 0$ allora alla prima giocata è risultato croce. Assumendo dunque che ogni lancio è indipendente da quello precedente, dopo la prima croce dobbiamo aggiungere un numero di lanci in media pari a $E[N]$ per ottenere testa.
Riscriviamo l'equazione:

$$E[N] = 1 \cdot p + (1-p)(1 + E[N]) \qquad \text{da cui} \qquad E[N] = \frac{1}{p}\,.$$

Ad esempio se $p = 0.8$ allora il numero medio di lanci per ottenere testa è 1.25; se $p = 0.05$ allora il numero medio è venti.

Osserviamo che N è una variabile aleatoria geometrica di parametro p con densità $p_N(n) = p(1-p)^{n-1}$. Senza ricorrere al valore atteso condizionato, possiamo calcolare

$$E[N] = \sum_{n=1}^{+\infty} n p_N(n) = \sum_{n=1}^{+\infty} n p (1-p)^{n-1} = \frac{1}{p}$$

in base alla proprietà delle variabili aleatorie geometriche.

3.3 Varianza condizionata

Prima di definire la varianza condizionata, vogliamo mostrare che si può usare il valore atteso condizionato per calcolare la varianza di una variabile aleatoria X. Sappiamo che $Var(X) = E[X^2] - (E[X])^2$, e quindi usiamo il condizionamento per ottenere $E[X]$ e $E[X^2]$. Vediamo subito un esempio.

Esempio 3.3.1. Calcoliamo la varianza di una variabile aleatoria geometrica. Supponiamo di eseguire una serie di prove indipendenti e che ciascuna prova abbia probabilità di successo p. Indichiamo con N la prova del primo successo. Vogliamo conoscere $Var(N) = E[N^2] - (E[N])^2$.
Sia Y variabile aleatoria così definita:

$$Y = \begin{cases} 1 & \text{se la prima prova è un successo} \\ 0 & \text{se la prima prova è un insuccesso} \end{cases}$$

Calcoliamo $E[N]$ e $E[N^2]$ usando la proprietà del valore atteso condizionato. La prima è già stata calcolata nell'esempio precedente ed è $E[N] = 1/p$. La seconda invece:

$$E[N^2] = E[E[N^2|Y]] = E[N^2|Y = 1]P\{Y = 1\} + E[N^2|Y = 0]P\{Y = 0\}$$

Ricordiamo $P\{Y=1\}=p$ e $P\{Y=0\}=1-p$.
Per i momenti secondi condizionati abbiamo:

$$E[N^2|Y=1]=1 \qquad E[N^2|Y=0]=E[(1+N)^2].$$

Infatti se la prima prova è un successo ($Y=1$), allora $N=1$ e dunque N^2=1. Se invece la prima prova è un insuccesso ($Y=0$), allora il numero totale di prove necessarie per ottenere il primo successo è uno più il numero di prove necessarie per ottenere il primo successo e cioè N. Dunque $N^2|Y=0$ è distribuita come $(1+N)^2$ e $E[N^2|Y=0]=E[(1+N)^2]$.
Calcoliamo infine:

$$E[N^2]=1\cdot p+E[(1+N)^2](1-p)=p+E(1+2N+N^2)(1-p)=$$

$$=p+(1+2E[N]+E[N^2])(1-p)=p+(1+\frac{2}{p}+E[N^2])-p-2-pE[N^2].$$

Dunque $E[N^2]=\frac{2-p}{p^2}$.
Calcoliamo infine

$$Var(N)=E[N^2]-(E[N])^2=\frac{2-p}{p^2}-(\frac{1}{p})^2=\frac{1-p}{p^2} \quad .$$

Un altro modo per determinare la varianza di una variabile aleatoria è quello di applicare la varianza condizionata. La **varianza condizionata** di X dato $Y=y$ è definita da

$$Var(X|Y=y)=E[(X-E[X|Y=y])^2|Y=y].$$

La varianza condizionata è definita allo stesso modo della varianza ordinaria con l'eccezione che tutte le probabilità sono condizionate all'evento $\{Y=y\}$.
Proviamo a espandere il quadrato:

$$E[(X-E[X|Y=y])^2|Y=y]=E[X^2-2XE[X|Y=y]+(E[X|Y=y])^2|Y=y]=$$

$$=E[X^2|Y=y]-2E[X|Y=y]E[X|Y=y]+(E[X|Y=y])^2=$$
$$=E[X^2|Y=y]-(E[X|Y=y])^2=Var(X|Y=y).$$

Dimostriamo ora la seguente formula:

Proposizione 3.3.1. *Per tutte le variabili aleatorie discrete X e Y, con X avente momento di ordine due finito, si ha:*

$$Var(X)=E[Var(X|Y)]+Var(E[X|Y]).$$

Sviluppiamo separatamente i due addendi a destra dell'uguaglianza:

$$E[Var(X|Y)] = E[E[X^2|Y] - (E[X|Y])^2] = E[E[X^2|Y]] - E[(E[X|Y])^2] =$$

$$= E[X^2] - E[(E[X|Y])^2].$$

L'altro:

$$Var(E[X|Y]) = E[(E[X|Y])^2] - (E[E[X|Y]])^2 = E[(E[X|Y])^2] - (E[X])^2.$$

Infine:

$$E[Var(X|Y)]+Var(E[X|Y]) = E[X^2]-E[(E[X|Y])^2]+E[(E[X|Y])^2]-(E[X])^2 =$$

$$= E[X^2] - (E[X])^2 = Var(X).$$

Facciamo subito un esempio di applicazione.

Esempio 3.3.2. Siano $X_1, X_2, ...$ variabili aleatorie indipendenti e identicamente distribuite con valore atteso m e varianza v. Assumiamo inoltre che esse siano indipendenti dalla variabile aleatoria intera non negativa N. Calcoliamo la varianza della variabile $S = \sum_{i=1}^{N} X_i$.
La variabile S è una somma aleatoria di variabile aleatorie indipendenti e identicamente distribuite e abbiamo calcolato il suo valore atteso nell'esempio 3.2.3: $E[S] = E[X]E[N]$, dove X ha la stessa distribuzione di una delle X_i.
Usando la formula troviamo: $Var(S) = E[Var(S|N)] + Var(E[S|N])$.

$$Var(S|N = n) = Var(\sum_{i=1}^{N} X_i|N = n) = Var(\sum_{i=1}^{n} X_i|N = n) = Var(\sum_{i=1}^{n} X_i) = nv$$

perché le X_i sono indipendenti da N. Concludiamo il primo calcolo:

$$Var(S|N) = Nv \qquad E[Var(S|N)] = E[Nv] = vE[N].$$

Calcoliamo ora:

$$E[S|N = n] = E[\sum_{i=1}^{N} X_i|N = n] = E[\sum_{i=1}^{n} X_i|N = n] =$$

$$= E[\sum_{i=1}^{n} X_i] = nm$$

perché le X_i sono indipendenti da N. Troviamo infine:

$$E[S|N] = Nm \qquad Var(E[S|N]) = Var(Nm) = m^2 Var(N).$$

Allora la varianza di S è

$$Var(S) = vE[N] + m^2 Var(N).$$

Supponiamo che N sia una variabile aleatoria di Poisson di parametro λ: $E[N] = \lambda$ e $Var(N) = \lambda$. Risulta infine:

$$Var(S) = v\lambda + m^2\lambda = \lambda E[X^2]$$

dove X è distribuita come le X_i.

3.4 Calcolare probabilità con il condizionamento

Usando l'approccio del condizionamento, possiamo calcolare le probabilità. Sia E un evento arbitrario e ricordiamo la variabile aleatoria indicatore dell'evento E, I_E:

$$I_E = \begin{cases} 1 & \text{se si verifica l'evento E} \\ 0 & \text{se non si verifica l'evento E} \end{cases}$$

Se calcoliamo il valore atteso di I_E, otteniamo:

$$E[I_E] = 1 \cdot P(E) + 0 \cdot (1 - P(E)) = P(E).$$

Se calcoliamo il valore atteso condizionato di I_E dato $Y = y$, dove Y è una variabile casuale qualsiasi, otteniamo:

$$E[I_E|Y = y] = 1 \cdot P\{I_E = 1|Y = y\} + 0 \cdot P\{I_E = 0|Y = y\} = P(E|Y = y).$$

E per la proprietà del valore atteso condizionato, cioè
$E[X] = \sum_y E[X|Y = y]P\{Y = y\}$, otteniamo anche il seguente risultato:

$$P(E) = \sum_y P(E|Y = y)P\{Y = y\} \ .$$

Vediamo degli esempi.

Esempio 3.4.1. Supponiamo che il numero di persone che frequenta una palestra in un giorno sia rappresentabile da una variabile aleatoria di Poisson con parametro λ. Supponiamo che ogni persona che frequenta sia,indipendentemente, femmina con probabilità p e maschio con probabilità $1 - p$. Determiniamo la probabilità congiunta che esattamente n donne e m uomini vadano in palestra

oggi.
Denotiamo con D il numero delle donne e con U il numero degli uomini che vanno in palestra oggi. Indichiamo con N il numero totale delle persone che vanno in palestra oggi: $N = D + U$. Applichiamo la relazione appena vista all'evento $\{D = n, U = m\}$:

$$P\{D = n, U = m\} = \sum_{i=0}^{+\infty} P\{D = n, U = m | N = i\} P\{N = i\}.$$

Osserviamo che se $i \neq n + m$, allora $P\{D = n, U = m | N = i\} = 0$. Quindi

$$P\{D = n, U = m\} = P\{D = n, U = m | N = n + m\} P\{N = n + m\}.$$

Essendo N una variabile di Poisson $P\{N = n + m\} = e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!}$.
Sappiamo che i visitatori sono in tutto $n+m$ e che ognuno di questi è indipendentemente donna con probabilià p. Segue che la probabilità condizionata che n di essi siano donne (e m siano uomini) è la probabilità binomiale che si verifichino n successi su $n + m$ prove.
Dunque $P\{D = n, U = m | N = n + m\} = \binom{n+m}{n} p^n (1 - p)^m$.
Calcoliamo infine:

$$P\{D = n, U = m\} = \binom{n+m}{n} p^n (1-p)^m e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!} =$$

$$= \frac{1}{n!m!} p^n (1-p)^m e^{-\lambda} e^{-\lambda p + \lambda p} \lambda^{n+m} = \frac{1}{n!m!} p^n (1-p)^m e^{-\lambda p} e^{-\lambda(1-p)} \lambda^{n+m} =$$

$$= e^{-\lambda p} \frac{(\lambda p)^n}{n!} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} \quad .$$

Osservando questo risultato viene il dubbio che D e U siano variabili indipendenti. Infatti il prodotto finale è costituito da due fattori: il primo dipende solo da n, il secondo solo da m. Verifichiamo.
Gli eventi $\{U = m\}$, al variare di $m = 0, 1, 2...$ sono disgiunti due a due. Pertanto posso scrivere $\{D = n\} = \bigcup_{m=0}^{+\infty} \{D = n, U = m\}$.
Calcoliamo le densità marginali:

$$P\{D = n\} = \sum_{m=0}^{+\infty} P\{D = n, U = m\} = \sum_{m=0}^{+\infty} e^{-\lambda p} \frac{(\lambda p)^n}{n!} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} =$$

$$= e^{-\lambda p} \frac{(\lambda p)^n}{n!} e^{-\lambda(1-p)} \sum_{m=0}^{+\infty} \frac{(\lambda(1-p))^m}{m!} = e^{-\lambda p} \frac{(\lambda p)^n}{n!} e^{-\lambda(1-p)} e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^n}{n!} \ .$$

Calcoliamo allo stesso modo l'altra. Stavolta abbiamo $\{U = m\} = \bigcup_{n=0}^{+\infty} \{D = n, U = m\}$ e

$$P\{U = m\} = \sum_{n=0}^{+\infty} P\{D = n, U = m\} = \sum_{n=0}^{+\infty} e^{-\lambda p} \frac{(\lambda p)^n}{n!} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} =$$

$$= e^{-\lambda p} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} \sum_{m=0}^{+\infty} \frac{(\lambda p)^n}{n!} = e^{-\lambda p} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} e^{\lambda p} =$$

$$= e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} \quad .$$

Per la definizione di variabili indipendenti, D e U sono variabili aleatorie indipendenti. In particolare sono variabili di Poisson di parametri rispettivamente λp e $\lambda(1-p)$.

Esempio 3.4.2. Ad una festa n uomini si tolgono i loro cappelli. I cappelli vengono messi in un scatolone e mescolati, quindi ogni uomo prende a caso un cappello. Abbiamo un successo se un uomo "pesca" proprio il suo cappello. Ci chiediamo quale sia la probabilità che non ci sia alcun successo.
Sia E l'evento "Non si verifica alcun successo", e siccome $P(E)$ dipende da n scriviamo $P_n = P(E)$. Chiamiamo M l'evento "il primo uomo prende il suo cappello". Con la formula delle probabilità totali scriviamo: $P_n = P(E|M)P(M) + P(E|\overline{M})P(\overline{M})$. Per come abbiamo definito gli eventi risulta $P(E|M) = 0$. Inoltre la probabilità che il primo uomo prenda il suo cappello è $P(M) = 1/n$, dunque $P(\overline{M}) = (n-1)/n$. Pertanto $P_n = P(E|\overline{M})\frac{n-1}{n}$
Ora $P(E|\overline{M})$ è la probabilità che non vi sia alcun successo quando $n-1$ uomini scelgono da un insieme di $n-1$ cappelli, che non contiene il cappello di uno dei signori (perché è stato preso dal primo). Chiamiamo C il cappello che ha preso il primo uomo, X il proprietario di C, D il cappello del primo uomo. L'evento $\{E|\overline{M}\}$ può accadere in due modi:

- $E1$: non si verifica alcun successo e X non prende D;
- $E2$: non si verifica alcun successo e X prende D.

Dunque $P(E|\overline{M}) = P(E1 \cup E2) = P(E1) + P(E2)$, perché $E1$ e $E2$ sono disgiunti. Calcoliamo separatamente $P(E1)$ e $P(E2)$. Chiamiamo M_1 l'evento "X prende D". Allora posso scrivere $P(E1) = P(E \cap \overline{M_1}) = P(E|\overline{M_1})P(\overline{M_1})$. Se consideriamo D come il cappello di X allora è $P(E_1) = P_{n-1}$. Infatti P_n

era risultata la probabilità che non si verificasse alcun successo su n e il primo uomo non scegliesse il suo cappello. P_{n-1} è invece la probabilità che non si verifichi alcun successo su $n-1$ e il primo uomo non scelga il suo cappello. Il secondo è invece $P(E2) = P(E \cap M_1) = P(E|M_1)P(M_1)$. $P(M_1)$ è la probabilità che X prenda D su $n-1$ cappelli e quindi $P(M_1) = 1/(n-1)$. $P(E|M_1)$ è la probabilità che non si verifichi alcun successo sapendo che X prende D. Ma se X prende D, i rimanenti $n-2$ signori devono scegliere un cappello tra un insieme di $n-2$. Ma ogni signore ha il proprio cappello in quell'insieme e quindi noi dobbiamo calcolare la probabilità che non si verifichi alcun successo su $n-2$. Avendo definito inizialmente P_n come la probabilità che non si verifichi alcun successo su n, allora $P(E|M_1) = P_{n-2}$.
Scriviamo infine:$P(E|\overline{M}) = P_{n-1} + \frac{1}{n-1} P_{n-2}$ e riscriviamo infine $P_n = P(E)$:

$$P_n = \frac{n-1}{n} P_{n-1} + \frac{1}{n} P_{n-2} \qquad \longrightarrow \qquad P_n - P_{n-1} = -\frac{1}{n}(P_{n-1} - P_{n-2})$$

Ora consideriamo il caso $n = 1$: abbiamo un solo uomo che deve scegliere un solo cappello, che è il suo. Non può non prendere il suo cappello. Pertanto $P_1 = 0$.
Consideriamo il caso $n = 2$: ci sono due uomini che devono scegliere tra due cappelli. La probabilità che non si verifichi alcun successo è la probabilità che il primo uomo prenda il cappello dell'altro e quindi $P_2 = 1/2$.
Ora che abbiamo determinato

$$P_1 = 0 \qquad e \qquad P_2 = \frac{1}{2}$$

possiamo determinare P_n per ogni $n \geq 3$, usando la relazione trovata sopra. Vediamo di trovare una forma più esplicita per calcolare P_n. Cominciamo col calcolare P_3 e P_4:

$$P_3 - P_2 = -\frac{1}{3} \cdot \frac{1}{2} = -\frac{1}{3!} \qquad \longrightarrow \qquad P_3 = \frac{1}{2!} - \frac{1}{3!}$$

$$P_4 - P_3 = -\frac{1}{4}\left(\frac{1}{2!} - \frac{1}{3!} - \frac{1}{2}\right) \qquad \longrightarrow \qquad P_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}$$

Ci viene il sospetto che la formula generale sia

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - ... + (-1)^n \frac{1}{n!}$$

Lo dimostriamo per induzione (nella seconda forma). Supponiamo che tale formula sia vera per tutti i casi $\leq n-1$. Mostriamo che è vera per n:

$$P_n = P_{n-1} - \frac{1}{n}(P_{n-1} - P_{n-2}) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - ... + (-1)^{n-1}\frac{1}{(n-1)!} +$$

$$-\frac{1}{n}(\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-...+(-1)^{n-2}\frac{1}{(n-2)!}+(-1)^{n-1}\frac{1}{(n-1)!}+$$
$$-\frac{1}{2!}+\frac{1}{3!}-\frac{1}{4!}+...-(-1)^{n-2}\frac{1}{(n-2)!})=$$
$$=\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-...+(-1)^{n-1}\frac{1}{(n-1)!}-\frac{1}{n}\frac{(-1)^{n-1}}{(n-1)!}=\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-...+(-1)^{n}\frac{1}{n!}$$

E quindi questa formula è vera per ogni numero naturale n.

Vogliamo calcolare infine qual è la probabilità di avere esattamente k successi, cioè k scelgono proprio il loro cappello. La chiamiamo $P(E_k)$. Prendiamo un gruppo fisso di k uomini. Il primo che pesca ha la probabilità di pescare il suo $1/n$, il secondo $1/(n-1)$, il terzo $1/(n-2)$,..., fino al k-esimo che ha probabilità di prendere il suo cappello $1/(n-(k-1))$. Stiamo usando sostanzialmente la formula della moltiplicazione. Dobbiamo moltiplicare tutti questi fattori per la probabilità che gli altri $n-k$ signori non prendano il loro cappello e cioè P_{n-k}. Ma abbiamo considerato un gruppo fissato di k persone. In realtà i gruppi di k persone che possiamo fare a partire da n sono in tutto $\binom{n}{k}$. In conclusione:

$$P(E_k)=\binom{n}{k}\frac{1}{n}\cdot\frac{1}{n-1}\cdot\frac{1}{n-2}\cdots\frac{1}{n-k+1}P_{n-k}=\frac{1}{k!}P_{n-k}$$

E per calcolare P_{n-k} uso la formula sopra.

Vediamo ora che l'equazione $P_n-P_{n-1}=-\frac{1}{n}(P_{n-1}-P_{n-2})$ può essere ottenuta in un altro modo.

Diciamo che la sequenza di individui distinti $i_1,i_2,...,i_k$ costituisce un ciclo di lunghezza k se i_1 sceglie il cappello di i_2, i_2 sceglie il cappello di i_3,...,i_{k-1} sceglie il cappello di i_k e infine i_k sceglie il cappello di i_1. Abbiamo un ciclo di lunghezza $k=1$ quando qualcuno sceglie il proprio cappello. Allora la probabilità che accada l'evento E si può calcolare così:

$$P(E)=P_n=\sum_{k=1}^{n}P(E|C=k)P\{C=k\}$$

dove C è la lunghezza del ciclo che contiene la persona 1. Determiniamo $P\{C=k\}$. Sia la persona 1 la prima persona che sceglie il cappello. Avremo $C=k$ se la prima persona non sceglie il suo cappello (e le possibilità sono $(n-1)/n$); la persona (che chiamiamo seconda persona) il cui cappello è stato scelto dalla prima non pesca il cappello della persona 1 (e le possibilità sono $(n-2)/(n-1)$); la persona (che chiamiamo terza persona) il cui cappello è stato scelto dalla seconda non pesca il cappello della persona 1 (e le possibilità sono $(n-3)/(n-2))$;...;la persona il cui cappello è stato scelto dalla $(k-1)$-esima pesca il cappello della persona 1 (e le possibilità sono $1/(n-(k-1)))$. Dunque

$$P\{C=k\} = \frac{n-1}{n}\frac{n-2}{n-1}\cdots\frac{n-k+1}{n-k+2}\frac{1}{n-k+1} = \frac{1}{n}.$$

Determiniamo ora $P(E|C=k)$. Se $k=1$ significa che una persona sceglie il suo cappello, quindi c'è un successo e $P(E|C=1)=0$. Se $k \neq 1$ significa che nessuna delle k persone sceglie il loro cappello. Dobbiamo calcolare la probabilità che nessuna delle $n-k$ persone rimanenti scelga il proprio cappello. Ma questa è P_{n-k}. Dunque

$$P(E|C=k) = P_{n-k} \quad \textit{per } \; k=2,3,...,n.$$

Infine

$$P_n = \frac{1}{n}\sum_{k=2}^{n} P_{n-k}.$$

Verifichiamo che otteniamo ancora l'equazione $P_n - P_{n-1} = -\frac{1}{n}(P_{n-1}-P_{n-2})$:

$$P_n - P_{n-1} = \frac{1}{n}\sum_{k=2}^{n} P_{n-k} - \frac{1}{n-1}\sum_{k=2}^{n-2} P_{n-k-1} =$$

$$= \frac{1}{n}P_{n-2} + (P_{n-3}+P_{n-4}+...+P_0)(\frac{1}{n}-\frac{1}{n-1}) =$$

$$= \frac{1}{n}P_{n-2} - \frac{1}{n(n-1)}(P_{n-3}+P_{n-4}+...+P_0) = \frac{1}{n}(P_{n-2}-P_{n-1}).$$

Abbiamo visto che $E[X] = \sum\limits_{y} E[X|Y=y]P\{Y=y\}$. Mostriamo una cosa analoga per $E[X|Y=y]$:

Proposizione 3.4.1. *Per tutte le variabili aleatorie discrete X, Y, W si ha*

$$E[X|Y=y] = \sum_{w} E[X|W=w, Y=y]P\{W=w|Y=y\}.$$

Infatti:

$$\sum_{w} E[X|W=w, Y=y]P\{W=w|Y=y\} =$$

$$= \sum_{x}\sum_{w} xP\{X=x|W=w, Y=y\}P\{W=w|Y=y\} =$$

$$= \sum_x x \sum_w \frac{P\{X = x, W = w, Y = y\}}{P\{W = w, Y = y\}} P\{W = w|Y = y\} =$$

$$= \sum_x x \sum_w \frac{P\{X = x, W = w, Y = y\}}{P\{W = w|Y = y\}P\{Y = y\}} P\{W = w|Y = y\} =$$

$$= \sum_x \frac{x}{P\{Y = y\}} \sum_w P\{X = x, Y = y, W = w\} =$$

$$= \sum_x \frac{xP\{X = x, Y = y\}}{P\{Y = y\}} = \sum_x xP\{X = x|Y = y\} = E[X|Y = y].$$

3.5 Il problema del collezionista di figurine

Nel campo della probabilità con il nome di "problema del collezionista di figurine" si intendono i problemi del tipo: completare una raccolta di figurine, una collezione di regalini che si possono trovare nelle confezioni di merendine o di altri prodotti. Quando compriamo figurine o merendine può succedere che il pezzo che troviamo faccia già parte della nostra collezione. Ci chiediamo quante figurine o scatole di merendine dobbiamo comprare in media per completare la raccolta o collezione.

Iniziamo con un caso semplice. Supponiamo di voler completare una raccolta di tre figurine. Indichiamo con X il numero di figurine che dovremo acquistare per completare la raccolta. Vogliamo conoscere dunque $E[X]$. Indichiamo poi con X_i il numero di figurine da acquistare per trovare la prima i-esima figurina diversa dalle figurine precedentemente trovate. Cioè si contano le figurine acquistate dopo che si è trovata la figurina $(i-1)$-esima. Sia inoltre p_i la probabilità di trovare la i-esima figurina della raccolta ad ogni singolo acquisto, supponendo che ogni acquisto sia indipendente dagli altri. Chiaramente $p_1 + p_2 + p_3 = 1$. Analizziamo:

- X_1 = numero di figurine che ci servono per trovare la prima figurina diversa dalle altre (ovviamente è uguale a 1 perché è la prima che acquistiamo!);
- X_2 = numero di figurine che acquistiamo per trovare una figurina diversa dalla prima, dopo che ho trovato la prima:
- X_3 = numero di figurine che acquistiamo per trovare la terza figurina che ci manca, dopo che ho trovato la seconda.

Ad esempio, $X_3 = 5$, vuol dire che abbiamo trovato quattro figurine uguali alla prima o alla seconda, e che al quinto acquisto abbiamo finalmente trovato

l'ultima figurina che completa la raccolta.
Dunque $X = X_1 + X_2 + X_3$ e per la linearità del valore atteso abbiamo $E[X] = E[X_1] + E[X_2] + E[X_3]$. Calcoliamo separatamente i tre valori attesi.
Il primo è facile: essendo $X_1 = 1$, si ha $E[X_1] = 1$.
Vediamo $E[X_2]$. Introduciamo la variabile aleatoria Y_m che indica quale figurina ho trovato all'm-esimo acquisto. Dunque Y_m assume il valore i con probabilità p_i, per $i = 1, 2, 3$. Per come abbiamo definito X_2, la variabile $X_2|Y_1 = i$ è il numero di figurine che dobbiamo acquistare per avere una figurina diversa da i, sapendo che al primo acquisto ho trovato la figurina i. Si tratta chiaramente di una variabile geometrica di parametro $(1 - p_i)$ e il suo valore atteso risulta: $E[X_2|Y_1 = i] = (1 - p_i)^{-1}$. Calcoliamo $E[X_2]$ usando la proprietà del valore atteso condizionato provata alla proposizione 3.2.1:

$$E[X_2] = \sum_{i=1}^{3} E[X_2|Y_1 = i]P\{Y_1 = i\} = \sum_{i=1}^{3} \frac{p_i}{1 - p_i} \quad .$$

Rimane $E[X_3]$. Condizioniamo X_3 con altre variabili cioè con quale figurine sono stare trovate come prima e come seconda: $X_3|Y_1 = i, Y_{1+X_2} = j, X_2 = k$, con $i \neq j$. È il numero di figurine che devo acquistare, dopo avere trovato la seconda, per avere una figurina diversa da i e da j, sapendo che al primo acquisto ho trovato i e all'acquisto $(k+1)$ ho trovato j. Si tratta di una variabile geometrica di parametro $(1 - p_i - p_j)$ e il suo valore atteso è: $E[X_3|Y_1 = i, Y_{1+X_2} = j, X_2 = k] = (1 - p_i - p_j)^{-1}$. Anche per calcolare $E[X_3]$ usiamo la proprietà del valore atteso condizionato provata alla proposizione 3.2.1:

$$E[X_3] = E[E[X_3|Y_1, Y_{1+X_2}, X_2]] =$$

$$= \sum_{k=1}^{+\infty} \sum_{i \neq j=1}^{3} E[X_3|Y_1 = i, Y_{1+X_2} = j, X_2 = k]P\{Y_1 = i, Y_{1+X_2} = j, X_2 = k\}$$

Il primo fattore è già stato determinato, per il secondo usiamo la formula della moltiplicazione:

$$P\{Y_1 = i, Y_{1+X_2} = j, X_2 = k\} =$$
$$= P\{Y_1 = i\}P\{X_2 = k|Y_1 = i\}P\{Y_{1+X_2} = j|Y_1 = i, X_2 = k\}$$

Calcoliamo separatamente le probabilità:

- $P\{Y_1 = i\} = p_i$
- $P\{X_2 = k|Y_1 = i\} = (1 - p_i)(1 - (1 - p_i))^{k-1} = (1 - p_i)p_i^{k-1}$ perché $X_2|Y_1 = i$ è una variabile geometrica di parametro $(1 - p_i)$

- $P\{Y_{1+X_2}=j|Y_1=i,X_2=k\}=O$ se $i=j$. Altrimenti, per $i\neq j$ si ha $P\{Y_{1+X_2}=j|Y_1=i,X_2=k\}=P\{Y_{1+k}=j|Y_{1+k}\neq i\}=$ $=\frac{P\{Y_{1+k}=j,Y_{1+k}\neq i\}}{P\{Y_{1+k}\neq i\}}=\frac{P\{Y_{1+k}=j\}}{P\{Y_{1+k}\neq i\}}=\frac{p_j}{1-p_i}$

Infine

$$E[X_3]=\sum_{k=1}^{+\infty}\sum_{i\neq j=1}^{3}\frac{1}{1-p_i-p_j}\frac{p_ip_j}{1-p_i}(1-p_i)p_i^{k-1}=$$

$$=\sum_{i\neq j=1}^{3}\frac{p_ip_j}{1-p_i-p_j}\sum_{k=1}^{+\infty}p_i^{k-1}=\sum_{i\neq j=1}^{3}\frac{p_ip_j}{(1-p_i-p_j)(1-p_i)}$$

Possiamo ora determinare $E[X]$:

$$E[X]=E[X_1]+E[X_2]+E[X_3]=1+\sum_{i=1}^{3}\frac{p_i}{1-p_i}+\sum_{i\neq j=1}^{3}\frac{p_ip_j}{(1-p_i-p_j)(1-p_i)}\ .$$

Generalizziamo a una raccolta di n figurine. Vogliamo determinare ora questo risultato:

$$E[X]=\sum_{i=1}^{n}E[X_i]$$

sapendo che ho probabilità p_i di trovare la figurina i, per $i=1,2,...,n$. Come prima $E[X_1]=1$. E per una variabile generale X_k:

$$E[X_k]=E[E[X_k|Y_1,Y_{1+X_2},X_2,Y_{1+X_2+X_3},X_3,...,Y_{1+X_2+...+X_{k-1}},X_{k-1}]=$$

$$=\sum_{h_2,...,h_{k-1}=1}^{+\infty}\sum_{i_1,...,i_{k-1}=1}^{n}E[X_k|Y_1=i_1,Y_{1+X_2}=i_2,X_2=h_2,Y_{1+X_2+X_3}=i_3,$$

$$X_3=h_3,...Y_{1+X_2+...+X_{k-1}}=i_{k-1},X_{k-1}=h_{k-1}]P\{Y_1=i_1,...,X_{k-1}=h_{k-1}\}$$

La variabile $X_k|Y_1=i_1,Y_{1+X_2}=i_2,X_2=h_2,Y_{1+X_2+X_3}=i_3,X_3=h_3,...,$ $Y_{1+X_2+...+X_{k-1}}=i_{k-1},X_{k-1}=h_{k-1}$ è il numero di figurine che devo acquistare per trovare la nuova k-esima figurina sapendo che al primo acquisto ho trovato i_1, poi, dopo altri h_2 acquisti ho trovato i_2 e dopo altri h_3 acquisti ho trovato i_3,... con gli i_t diversi tutti tra loro. È una variabile geometrica di parametro $1-p_{i_1}-p_{i_2}-...-p_{i_{k-1}}$ e quindi il suo valore atteso è $(1-p_{i_1}-p_{i_2}-...-p_{i_{k-1}})^{-1}$. Possiamo scrivere l'ultimo fattore, per la formula di moltiplicazione, come:

$$P\{Y_1=i_1\}P\{X_2=h_2|Y_1=i_1\}P\{Y_{1+X_2}=i_2|Y_1=i_1,X_2=h_2\}\cdots$$

$$\cdots P\{Y_{1+X_2+...+X_{k-1}}=i_{k-1}|Y_1=i_1,Y_{1+X_2}=i_2,...X_{k-1}=h_{k-1}\}$$

Calcoliamo i fattori separatamente:

- $P\{Y_1 = i_1\} = p_{i_1}$;
- $P\{X_2 = h_2 | Y_1 = i_1\} = (1 - p_{i_1})(1 - (1 - p_{i_1}))^{h_2 - 1}$, perché, come abbiamo già spiegato, $X_2 | Y_1 = i_1$ è variabile geometrica di parametro $1 - p_{i_1}$;
- $P\{Y_{1+X_2} = i_2 | Y_1 = i_1, X_2 = h_2\} = P\{Y_{1+h_2} = i_2 | Y_{1+h_2} \neq i_1\} =$
$= \frac{P\{Y_{1+h_2} = i_2, Y_{1+h_2} \neq i_1\}}{P\{Y_{1+h_2} \neq i_1\}} = \frac{p_{i_2}}{1 - p_{i_1}}$, con $i_1 \neq i_2$;

Osservazione: si noti che Y_1 e Y_k sono indipendenti $\forall k > 1$, ma non lo sono più quando condizioniamo al fatto che $X_2 = k$.

- $P\{1 + X_2 + X_3 = h_3 | Y_1 = i_1, Y_{1+X_2} = i_2, X_2 = h_2\} = (1 - p_{i_1} - p_{i_2})(1 - (1 - p_{i_1} - p_{i_2}))^{h_3 - 1}$. La spiegazione è data sopra, quando parliamo della variabile $X_k | Y_1 = i_1, Y_{1+X_2} = i_2, ...$. In questo caso $k = 3$;
- $P\{Y_{1+X_2+X_3} = i_3 | Y_1 = i_1, Y_{1+X_2} = i_2, X_2 = h_2, X_3 = h_3\} = P\{Y_{1+h_2+h_3} = i_3 | Y_{1+h_2+h_3} \neq i_1, Y_{1+h_2+h_3} \neq i_2\} =$
$= \frac{P\{Y_{1+h_2+h_3} = i_3, Y_{1+h_2+h_3} \neq i_1, Y_{1+h_2+h_3} \neq i_2\}}{P\{Y_{1+h_2+h_3} \neq i_1, Y_{1+h_2+h_3} \neq i_2\}} = \frac{p_{i_3}}{1 - p_{i_1} - p_{i_2}}$, con gli i_t diversi tra loro;
- ...
- $P\{X_{k-1} = h_{k-1} | Y_1 = i_1, Y_{1+X_2} = i_2, ..., X_{k-2} = h_{k-2}\} = (1 - p_{i_1} - p_{i_2} - ... - p_{i_{k-2}})(1 - (1 - p_{i_1} - p_{i_2} - ... - p_{i_{k-2}}))^{h_{k-1} - 1}$
- $P\{Y_{1+X_2+...+X_{k-1}} = i_{k-1} | Y_1 = i_1, Y_{1+X_2} = i_2, ... X_{k-1} = h_{k-1}\} =$
$= P\{Y_{1+...+h_{k-1}} = i_{k-1} | Y_{1+...+h_{k-1}} \neq i_1, ..., Y_{1+...+h_{k-1}} \neq i_{k-2}\} =$
$= \frac{P\{Y_{1+...+h_{k-1}} = i_{k-1}, Y_{1+...+h_{k-1}} \neq i_1, ..., Y_{1+...+h_{k-1}} \neq i_{k-2}\}}{P\{Y_{1+...+h_{k-1}} \neq i_1, ..., Y_{1+...+h_{k-1}} \neq i_{k-2}\}} =$
$= \frac{p_{i_{k-1}}}{1 - p_{i_1} - p_{i_2} - ... - p_{i_{k-2}}}$, con gli i_t diversi tra loro.

Infine mettiamo insieme tutti i fattori, e per semplificare la scrittura poniamo:

$$p(i_1) = 1 - p_{i_1} \ \ (= p_1 + ... + \widehat{p}_{i_1} + ... + p_n)$$

$$p(i_1, i_2) = 1 - p_{i_1} - p_{i_2} \ \ (= p_1 + ... + \widehat{p}_{i_1} + ... + \widehat{p}_{i_2} + ... p_n)$$

(con $\widehat{p}_{i_t}$ intendiamo dire che togliamo dalla somma p_{i_t})

$$\vdots$$

$$p(i_1, i_2, ..., i_{k-1}) = 1 - p_{i_1} - p_{i_2} - ... - p_{i_{k-1}}$$

$$E[X_k] = \sum_{h_2,\dots,h_{k-1}=1}^{+\infty} \sum_{i_1,\dots,i_{k-1}=1}^{n} \frac{p_{i_1}p_{i_2}\cdots p_{i_{k-1}}p(i_1)p(i_1,i_2)\cdots p(i_1,i_2,\dots,i_{k-2})}{p(i_1)p(i_1,i_2)\cdots p(i_1,i_2,\dots,i_{k-1})}\cdot$$

$$\cdot p_{i_1}^{h_2-1}(p_{i_1}+p_{i_2})^{h_3-1}\cdots(p_{i_1}+p_{i_2}+\dots+p_{i_{k-2}})^{h_{k-1}} =$$

$$= \sum_{i_1,\dots,i_{k-1}=1}^{n} \frac{p_{i_1}\cdots p_{i_{k-1}}}{p(i_1,\dots,i_{k-1})} \sum_{h_2,\dots,h_{k-1}=1}^{+\infty} p_{i_1}^{h_2-1}(p_{i_1}+p_{i_2})^{h_3-1}\cdots(p_{i_1}+\dots+p_{i_{k-2}})^{h_{k-1}} =$$

$$= \sum_{i_1,\dots,i_{k-1}=1}^{n} \frac{p_{i_1}\cdots p_{i_{k-1}}}{p(i_1,\dots,i_{k-1})} \sum_{h_2=1}^{+\infty} p_{i_1}^{h_2-1} \sum_{h_3=1}^{+\infty}(p_{i_1}+p_{i_2})^{h_3-1}\cdots$$

$$\cdots \sum_{h_{k-1}=1}^{+\infty}(p_{i_1}+\dots+p_{i_{k-2}})^{h_{k-1}} = \sum_{i_1,\dots,i_{k-1}=1}^{n} \frac{p_{i_1}p_{i_2}\cdots p_{i_{k-1}}}{p(i_1)p(i_1,i_2)\cdots p(i_1,i_2,\dots,i_{k-1})}$$

Otteniamo perciò l'espressione di $E[X]$:

$$E[X] = \sum_{k=1}^{n} E[X_k] =$$

$$= 1 + \sum_{i_1=1}^{n} \frac{p_{i_1}}{p(i_1)} + \sum_{i_1,i_2=1}^{n} \frac{p_{i_1}p_{i_2}}{p(i_1)p(i_1,i_2)} + \dots + \sum_{i_1,\dots,i_{n-1}=1}^{n} \frac{p_{i_1}\cdots p_{i_{n-1}}}{p(i_1)p(i_1,i_2)\cdots p(i_1,\dots,i_{n-1})}$$

Ricordiamo che gli i_t sono tutti diversi tra di loro.

Il valore $E[X]$ trovato qui può essere espresso in un altro modo. Si può infatti dimostrare che vale la seguente relazione [1]:

siano di nuovo $p_i > 0$, per $i = 1, 2, \dots, n$, reali, le probabilità di acquistare ad ogni singolo caso le figurine e tali che

$$p_1 + p_2 + \dots + p_n = 1.$$

Il numero atteso di figurine che si devono comprare per concludere la raccolta vale:

$$\sum_{1\le i\le n} \frac{1}{p_i} - \sum_{1\le i<j\le n} \frac{1}{p_i+p_j} + \sum_{1\le i<j<k\le n} \frac{1}{p_i+p_j+p_k} - \dots (-1)^{n+1} \frac{1}{p_1+p_2+\dots+p_n}$$

Da quanto visto prima, vale la seguente interessante identità combinatoria:

$$\sum_{1\le i\le n} \frac{1}{p_i} - \sum_{1\le i<j\le n} \frac{1}{p_i+p_j} + \sum_{1\le i<j<k\le n} \frac{1}{p_i+p_j+p_k} - \dots (-1)^{n+1} \frac{1}{p_1+p_2+\dots+p_n} =$$

[1] Vedi **S.M. Ross**, *A First Course in Probability, 6th Edition*, Prentice Hall, Upper Saddle River (USA), 2002, pag.324

$$= 1+\sum_{i=1}^{n}\frac{p_i}{p(i)}+\sum_{i\neq j=1}^{n}\frac{p_ip_j}{p(i)p(i,j)}+...+\sum_{i_1\neq ...\neq i_{n-1}=1}^{n}\frac{p_{i_1}p_{i_2}\cdots p_{i_{n-1}}}{p(i_1)p(i_1,i_2)\cdots p(i_1,i_2,...,i_{n-1})}$$

Il caso $n=2$ e $n=3$ si dimostra facilmente tramite operazioni algebriche. Per $n>3$ è troppo complicato procedere allo stesso modo e non sembra immediato poterlo dimostrare algebricamente. Vediamo i due casi semplici sopra citati.

Caso n=2 Innanzitutto abbiamo $p_1+p_2=1$ e vogliamo dimostrare che è vera la seguente uguaglianza:

$$\frac{1}{p_1}+\frac{1}{p_2}-\frac{1}{p_1+p_2}=1+\frac{p_1}{1-p_1}+\frac{p_2}{1-p_2}$$

Partiamo dall'espressione a destra dell'uguale:

$$1+\frac{p_1}{1-p_1}+\frac{p_2}{1-p_2}=1+\frac{p_1}{p_2}+\frac{p_2}{p_1}=\frac{p_1p_2+p_1^2+p_2^2}{p_1p_2}=\frac{(p_1+p_2)^2-p_1p_2}{p_1p_2}=$$

$$=\frac{1}{p_1p_2}-1=\frac{1}{p_1}+\frac{1}{p_2}-1=\frac{1}{p_1}+\frac{1}{p_2}-\frac{1}{p_1+p_2}$$

Ed è dimostrato.

Caso n=3 Adesso abbiamo $p_1+p_2+p_3=1$ e vogliamo dimostrare che è vera la seguente uguaglianza:

$$\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}-\frac{1}{p_1+p_2}-\frac{1}{p_1+p_3}-\frac{1}{p_2+p_3}+\frac{1}{p_1+p_2+p_3}=$$

$$=1+\frac{p_1}{1-p_1}+\frac{p_2}{1-p_2}+\frac{p_3}{1-p_3}+\frac{p_1p_2}{(1-p_1)(1-p_1-p_2)}+\frac{p_1p_2}{(1-p_2)(1-p_1-p_2)}+$$

$$+\frac{p_1p_3}{(1-p_1)(1-p_1-p_3)}+\frac{p_1p_3}{(1-p_3)(1-p_1-p_3)}+\frac{p_2p_3}{(1-p_2)(1-p_2-p_3)}+$$

$$+\frac{p_2p_3}{(1-p_3)(1-p_2-p_3)}$$

Aggiungiamo -1 sia destra che a sinistra dell'uguale e partiamo dall'espressione a destra dell'uguale, riscrivendo i denominatori, ricordandoci che $p_1+p_2+p_3=1$:

$$\frac{p_1}{p_2+p_3}+\frac{p_2}{p_1+p_3}+\frac{p_3}{p_1+p_2}+\frac{p_1p_2}{p_3(p_2+p_3)}+\frac{p_1p_2}{p_3(p_1+p_3)}+\frac{p_1p_3}{p_2(p_2+p_3)}+$$

$$+\frac{p_1p_3}{p_2(p_1+p_2)}+\frac{p_2p_3}{p_1(p_1+p_3)}+\frac{p_2p_3}{p_1(p_1+p_2)}=$$

$$= \frac{p_1}{p_2+p_3}(1+\frac{p_2}{p_3}+\frac{p_3}{p_2}) + \frac{p_2}{p_1+p_3}(1+\frac{p_1}{p_3}+\frac{p_3}{p_1}) + \frac{p_3}{p_1+p_2}(1+\frac{p_1}{p_2}+\frac{p_2}{p_1}) =$$

$$= \frac{p_1}{p_2+p_3}(\frac{p_2p_3+p_2^2+p_3^2}{p_2p_3}) + \frac{p_2}{p_1+p_3}(\frac{p_1p_3+p_1^2+p_3^2}{p_1p_3}) + \frac{p_3}{p_1+p_2}(\frac{p_1p_2+p_1^2+p_2^2}{p_1p_2}) =$$

$$= \frac{p_1}{p_2+p_3}(\frac{p_2(1-p_1-p_2)+p_2^2+p_3^2}{p_2p_3}) + \frac{p_2}{p_1+p_3}(\frac{(1-p_2-p_3)p_3+p_1^2+p_3^2}{p_1p_3}) +$$

$$+ \frac{p_3}{p_1+p_2}(\frac{p_1(1-p_1-p_3)+p_1^2+p_2^2}{p_1p_2}) =$$

$$= \frac{p_1}{p_2+p_3}(\frac{p_2-p_1p_2+p_3^2}{p_2p_3}) + \frac{p_2}{p_1+p_3}(\frac{p_3-p_2p_3+p_1^2}{p_1p_3}) + \frac{p_3}{p_1+p_2}(\frac{p_1-p_1p_3+p_2^2}{p_1p_2}) =$$

$$= \frac{p_1}{p_2+p_3}(\frac{1-p_1}{p_3}) + \frac{p_1p_3}{p_2(p_2+p_3)} + \frac{p_2}{p_1+p_3}(\frac{1-p_2}{p_1}) + \frac{p_1p_2}{p_3(p_1+p_3)} +$$

$$+ \frac{p_3}{p_1+p_2}(\frac{1-p_3}{p_2}) + \frac{p_2p_3}{p_1(p_1+p_2)} =$$

$$= \frac{p_1}{p_3}(1+\frac{p_2}{p_1+p_3}) + \frac{p_2}{p_1}(1+\frac{p_3}{p_1+p_2}) + \frac{p_3}{p_2}(1+\frac{p_1}{p_2+p_3}) =$$

$$= \frac{p_1}{p_3}(\frac{1}{p_1+p_3}) + \frac{p_2}{p_1}(\frac{1}{p_1+p_2}) + \frac{p_3}{p_2}(\frac{1}{p_2+p_3}) =$$

$$= \frac{1-p_2-p_3}{p_3(p_1+p_3)} + \frac{1-p_1-p_3}{p_1(p_1+p_2)} + \frac{1-p_1-p_2}{p_2(p_2+p_3)} =$$

$$= \frac{1}{p_3(p_1+p_3)} - \frac{p_2}{p_3(p_1+p_3)} - \frac{1}{p_1+p_3} + \frac{1}{p_1(p_1+p_2)} - \frac{1}{p_1+p_2} - \frac{p_3}{p_1(p_1+p_2)} +$$

$$+ \frac{1}{p_2(p_2+p_3)} - \frac{p_1}{p_2(p_2+p_3)} - \frac{1}{p_2+p_3} =$$

$$= -(\frac{1}{p_1+p_3} + \frac{1}{p_1+p_2} + \frac{1}{p_2+p_3}) + \frac{1}{p_3(p_1+p_3)}(1-p_2) +$$

$$+ \frac{1}{p_1(p_1+p_2)}(1-p_3) + \frac{1}{p_2(p_2+p_3)}(1-p_1) =$$

$$= \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} - \frac{1}{p_1+p_2} - \frac{1}{p_1+p_3} - \frac{1}{p_2+p_3}$$

Siamo riusciti a dimostrare anche questo caso.

Capitolo 4

Spazi di probabilità generali

Gli spazi di probabilità discreti si rivelano insufficienti per descrivere certi esperimenti. Vediamo subito un esempio.

Esempio 4.0.1. Consideriamo l'esperimento che consiste nel lanciare un numero n di volte una moneta equilibrata. I risultati del nostro esperimento, cioè gli elementi di Ω, saranno sequenze di simboli $0, 1$ (1 significa che è uscito testa, 0 croce) :

$$\Omega = \{\omega : \omega = (a_1, a_2, ..., a_n), a_i = 0, 1\}.$$

Allora Ω ha esattamente 2^n elementi e la probabilità che esca la sequenza ω è $P(\{\omega\}) = (1/2)^n$. Infine $P(\Omega) = \sum_{\omega \in \Omega} P(\{\omega\}) = 1$. Ora consideriamo l'esperimento che consiste nel lanciare un numero infinito di volte una moneta equilibrata. Analogamente a quanto fatto prima, descriviamo nel seguente modo Ω:

$$\Omega = \{\omega : \omega = (a_1, a_2, ...), a_i = 0, 1\}.$$

Quanti sono gli elementi di Ω? Osserviamo che ogni numero $a \in [0, 1)$ ha un'unica rappresentazione binaria (contenente un numero infinito di a_n uguale a zero) del tipo

$$a = \frac{a_1}{2} + \frac{a_2}{2^2} + ... + \frac{a_n}{2^n} + ... \qquad (a_n = 0, 1).$$

I punti dell'intervallo $[0, 1)$ sono dunque in corrispondenza biunivoca con le sequenze di $0, 1$ e quindi con gli elementi di Ω. Pertanto Ω è non numerabile. Dimostriamo che $P(\{\omega\}) = 0 \ \forall \omega \in \Omega$. Fissiamo $\omega = (\omega_1, \omega_2, \omega_3, ...)$. Definiti gli eventi
$A_1 = \{x \in \Omega : x_1 = \omega_1\}$
$A_2 = \{x \in \Omega : x_1 = \omega_1, x_2 = \omega_2\}$...

$A_n = \{x \in \Omega : x_1 = \omega_1, ..., x_n = \omega_n\}...$
da quanto fatto prima abbiamo: $P(A_1) = \frac{1}{2}, P(A_2) = \frac{1}{2^2}, ..., P(A_n) = \frac{1}{2^n}, ...$
Per la proprietà della probabilità, se P è una probabilità definita su Ω, avremo $P(\{\omega\}) = \lim_{n\to+\infty} P(A_n) = \lim_{n\to+\infty} \frac{1}{2^n} = 0$. È quindi chiaro che non possiamo definire la probabilità che accada un certo evento A con la formula $P(A) = \sum_{\omega\in A} P(\{\omega\})$.

Quanto visto suggerisce che dobbiamo costruire modelli che assegnino probabilità non ai singoli elementi di Ω, ma a sottoinsiemi di Ω. Da qui inizia l'associazione del concetto di probabilità con il concetto di misura di insiemi. In questo capitolo vedremo definizioni e teoremi della teoria della misura che ci servono per la costruzione di modelli di probabilità e in particolare per trattare il valore atteso condizionato.

4.1 σ-algebre e misure di probabilità

Sia Ω un insieme di elementi. Una famiglia $\mathcal{F}$ di sottoinsiemi di Ω è una **σ-algebra** se soddisfa alle seguenti condizioni:

1. $\Omega \in \mathcal{F}$;
2. $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$;
3. $A_n \in \mathcal{F} \;\; n = 1, 2, ... \Rightarrow \bigcup_n A_n \in \mathcal{F}$ e $\bigcap_n A_n \in \mathcal{F}$.

Dato un insieme Ω, i più semplici esempi di σ-algebre sono i seguenti:

- σ-algebra banale: $\mathcal{F}_* = \{\varnothing, \Omega\}$;
- σ-algebra generata da $A \subseteq \Omega$: $\mathcal{F}_A = \{A, \overline{A}, \varnothing, \Omega\}$;
- σ-algebra di tutte le parti di Ω: $\mathcal{F}^* = \mathcal{P}(\Omega)$.

Un'applicazione $P : \mathcal{F} \longmapsto [0, +\infty]$ si dice **misura di probabilità o probabilità** se soddisfa alle seguenti condizioni:

1. $P(\Omega) = 1 < +\infty$;
2. se $(A_n)_{n\in\mathbb{N}}$ è una successione di elementi di $\mathcal{F}$ a due a due disgiunti risulta
$$P(\bigcup_n A_n) = \sum_n P(A_n).$$

Per la seconda condizione si dice che la probabilità è una misura σ-additiva, e per la prima che è una misura finita.
La probabilità gode delle seguenti proprietà [1]:

1. $P(\varnothing) = 0$;
2. se $A, B \in \mathcal{F}$, allora $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
3. se $A, B \in \mathcal{F}$, allora $P(A \backslash B) = P(A) - (A \cap B)$;
4. se $A, B \in \mathcal{F}$ e $B \subseteq A$, allora $P(B) \leq P(A)$;
5. se $A_n \in \mathcal{F}$, $n = 1, 2, ...$, allora $P(A_1 \cup A_2 \cup ...) \leq P(A_1) + P(A_2) + ...$.

Il prossimo teorema stabilisce alcune condizioni equivalenti affinchè una misura di probabilità finitamente additiva $P : \mathcal{F} \longmapsto [0, +\infty]$ (che soddisfa cioè alle seguenti due condizioni: $1) A, B \in \mathcal{F},\ A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$; $2) P(\Omega) = 1 < +\infty$) sia σ-additiva.

Proposizione 4.1.1. [2] *Sia $P : \mathcal{F} \longmapsto [0, +\infty]$ una misura di probabilità finitamente additiva . Allora le seguenti quattro condizioni sono equivalenti:*

1. *P è σ-additiva (cioè P è una probabilità);*
2. *P è continua dal basso, cioè data una successione di eventi $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{F}$ tale che $A_n \subseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, risulta*
$$\lim_{n \to +\infty} P(A_n) = P(\bigcup_{n=1}^{+\infty} A_n);$$
3. *P è continua dall'alto, cioè data una successione di eventi $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{F}$ tale che $A_n \supseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, risulta*
$$\lim_{n \to +\infty} P(A_n) = P(\bigcap_{n=1}^{+\infty} A_n);$$
4. *P è continua nell'origine, cioè data una successione di eventi $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{F}$ tale che $A_n \supseteq A_{n+1}$ per ogni $n \in \mathbb{N}$ e $\bigcap_n A_n = \varnothing$, risulta*
$$\lim_{n \to +\infty} P(A_n) = 0.$$

[1] Per le dimostrazioni vedi **N. Cufaro Petroni**, *Lezioni di Calcolo delle Probabilità*, Edizioni dal Sud, Modugno(Bari), 1996, pag.129.

[2] Per le dimostrazioni vedi **A.N. Shiryayev**, *Probability*, Springer-Verlag, USA, 1984, pagg. 132-136.

Chiamiamo **spazio di probabilità** una terna ordinata $(\Omega, \mathcal{F}, P)$ dove Ω è un insieme di elementi ω detto **spazio campionario o degli eventi elementari**, $\mathcal{F}$ è una σ-algebra di sottoinsiemi (eventi) di Ω, P è una probabilità su $\mathcal{F}$.

4.2 La σ-algebra di Borel

Consideriamo ora uno dei più importanti esempi di σ-algebre che si incontrano nella teoria della probabilità.
Sia $\mathcal{A} = \{A\}$ una famiglia di sottoinsiemi di Ω. La più piccola σ-algebra che contiene $\mathcal{A}$ è detta la σ-algebra generata da $\mathcal{A}$ e si indica con $\mathcal{F}(\mathcal{A})$. Possiamo anche dire che $\mathcal{F}(\mathcal{A})$ è l'intersezione di tutte le σ-algebre $\mathcal{G}$ che contengono $\mathcal{A}$. Ad esempio la σ-algebra di tutte le parti di Ω contiene $\mathcal{A}$.
Assumiamo ora $\Omega = \mathbb{R}$. Chiamiamo **σ-algebra di Borel di** $\mathbb{R}$ la σ-algebra generata dagli aperti o equivalentemente dai chiusi di $\mathbb{R}$ e i suoi elementi sono detti **boreliani**. In genere è indicata con $\mathcal{B}(\mathbb{R})$. Siano $a, b \in \mathbb{R}$, $a < b$ e consideriamo le seguenti famiglie di sottoinsiemi:

1. $\mathcal{A}_1$ è la collezione di intervalli della forma (a, b);
2. $\mathcal{A}_2$ è la collezione di intervalli della forma $[a, b)$;
3. $\mathcal{A}_3$ è la collezione di intervalli della forma $(a, b]$;
4. $\mathcal{A}_4$ è la collezione di intervalli della forma $[a, b]$;
5. $\mathcal{A}_5$ è la collezione di intervalli della forma $(-\infty, a)$;
6. $\mathcal{A}_6$ è la collezione di intervalli della forma $(a, +\infty)$;
7. $\mathcal{A}_7$ è la collezione degli aperti di $\mathbb{R}$;
8. $\mathcal{A}_8$ è la collezione dei chiusi di $\mathbb{R}$.

Proposizione 4.2.1. $\mathcal{F}(\mathcal{A}_1) = \mathcal{F}(\mathcal{A}_2) = ... = \mathcal{F}(\mathcal{A}_8) = \mathcal{B}(\mathbb{R})$. *Vale a dire che la σ-algebra dei boreliani di $\mathbb{R}$ è generata da una qualunque delle famiglie di insiemi descritte sopra.*

Proviamo ad esempio che $\mathcal{F}(\mathcal{A}_5) = \mathcal{F}(\mathcal{A}_1)$. Le altre uguaglianze si dimostrano allo stesso modo.
Innanzitutto per ogni $a < b$ abbiamo $(-\infty, b) \backslash (-\infty, a) = [a, b) \in \mathcal{F}(\mathcal{A}_5)$. Prendiamo inoltre un intervallo (a, b) e una successione decrescente $(a_n)_{n \in \mathbb{N}}$ che converge ad a (ad esempio quella in cui ogni termine a_n si descrive come

$a_n = a + 1/n$). Allora $\bigcup_n [a_n, b) = (a, b)$ e anche $\bigcup_n [a_n, b) \in \mathcal{F}(\mathcal{A}_5)$. Quindi $(a, b) \in \mathcal{F}(\mathcal{A}_5)$ e in conclusione $\mathcal{F}(\mathcal{A}_1) \subseteq \mathcal{F}(\mathcal{A}_5)$. D'altra parte essendo $\bigcup_n [a_n, b) \in \mathcal{F}(\mathcal{A}_5)$, e che $\bigcup_n [a_n, b) = (a, b) \in \mathcal{F}(\mathcal{A}_1)$ segue che $\bigcup_n [a_n, b) \in \mathcal{F}(\mathcal{A}_1)$ e quindi $\mathcal{F}(\mathcal{A}_5) \subseteq \mathcal{F}(\mathcal{A}_1)$.
Dalla definizione di boreliano abbiamo che $\mathcal{F}(\mathcal{A}_7) = \mathcal{B}(\mathbb{R})$.
Se consideriamo $\mathcal{F}(\mathcal{A}_6)$, questa contiene gli intervalli inferiormente semiaperti, essendo $(a, b] = (a, +\infty) \setminus (b, +\infty)$. Prendiamo una successione $(b_n)_{n \in \mathbb{N}}$ crescente che converge a b (ad esempio quella in cui il termine b_n è definito come $b_n = b - 1/n$). Allora $\mathcal{F}(\mathcal{A}_6)$ contiene anche $\bigcup_n (a, b_n] = (a, b)$, cioè tutti gli intervalli aperti. Ora la σ-algebra generata dagli intervalli aperti contiene tutti gli aperti perché ogni aperto di $\mathbb{R}$ è unione numerabile di intervalli. Pertanto coincide con la σ-algebra generata dagli aperti, cioè $\mathcal{B}(\mathbb{R})$.

Al posto di $\mathbb{R}$ prendiamo un generico spazio topologico Ω, e chiamiamo $\mathcal{A}$ la famiglia degli aperti di Ω. La σ-algebra $\mathcal{B}(\Omega) = \mathcal{F}(\mathcal{A})$ generata da $\mathcal{A}$ è detta **σ-algebra di Borel di** Ω.

Se $P : \mathcal{B}(\Omega) \longmapsto [0, +\infty]$ è una probabilità, allora $(\Omega, \mathcal{B}(\Omega), P)$ è uno spazio di probabilità.

4.3 Probabilità condizionata e indipendenza

Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità. Dati due eventi $A, B \in \mathcal{F}$ con $P(B) \neq 0$, la **probabilità condizionata** di A rispetto a B è definita da

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

I teoremi che abbiamo visto in 2.3 valgono anche nel caso generale. Due eventi $A, B \in \mathcal{F}$ si dicono **eventi indipendenti** se

$$P(A \cap B) = P(A)P(B) \,.$$

In generale diciamo che n eventi $A_1, A_2, ..., A_n \in \mathcal{F}$ si dicono indipendenti se per ogni $k = 1, ..., n$ e per ogni insieme di indici $\{i_1, ..., i_k\}$ con $1 \leq i_1 < i_2 < ... < i_k \leq n$ risulta

$$P(A_1 \cap A_2 \cap ...A_n) = P(A_1)P(A_2) \cdots P(A_n).$$

4.4 Le variabili aleatorie

Dati uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e uno spazio probabilizzabile [3] $(E, \mathcal{E})$, una funzione $\xi : \Omega \longmapsto E$ si dice **variabile aleatoria** (o $\mathcal{F}$-misurabile) se

$$\xi^{-1}(B) = \{\xi \in B\} = \{\omega \in \Omega : \xi(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{E}.$$

Spesso si scrive $\xi : (\Omega, \mathcal{F}) \longmapsto (E, \mathcal{E})$ per indicare le σ-algebre rispetto alle quali la variabile aleatoria è misurabile.
Scegliamo $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $n \in \mathbb{N}$.
In particolare si dice che ξ è una **variabile aleatoria scalare** se $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$; si dice che è una **variabile aleatoria vettoriale o vettore aleatorio** se $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $n \in \mathbb{N}, n > 1$. In questo ultimo caso scriviamo $\xi = (\xi_1, ..., \xi_n)$ e le ξ_i sono variabili aleatorie scalari.

Chiamiamo **distribuzione o legge di probabilità** della variabile aleatoria scalare ξ la misura di probabilità P_ξ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definita da

$$P_\xi(B) = P\{\omega : \xi(\omega) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}).$$

Chiamiamo **funzione di distribuzione di** ξ la funzione

$$F_\xi(x) = P_\xi(-\infty, x] = P\{\xi \leq x\} = P\{\omega \in \Omega : \xi(\omega) \leq x\}, \quad x \in \mathbb{R}.$$

Per una variabile aleatoria discreta la misura P_ξ è concentrata su un insieme al più numerabile e può essere rappresentata nella seguente forma

$$P_\xi(B) = \sum_{\{k : x_k \in B\}} p(x_k) \qquad p(x_k) = P\{\xi = x_k\} = \Delta F_\xi(x_k).$$

Le proprietà di $F_\xi(x)$ descritte nel caso discreto, valgono anche nel caso generale.

Sia $\xi = (\xi_k)_{k=1,...,n}$ un vettore aleatorio.
Chiamiamo **distribuzione di probabilità** (delle componenti) di ξ la seguente funzione definita su $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$:

$$P_\xi(B) = P(\xi^{-1}(B)) = P\{\omega : \xi(\omega) \in B\}$$

dove B è un generico elemento di $\mathcal{B}(\mathbb{R}^n)$.
Chiamiamo **distribuzioni di probabilità marginali** (delle componenti) di ξ le seguenti funzioni definite su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$P_{\xi_k}(A) = P(\xi_k^{-1}(A)) = P\{\omega : \xi_k(\omega) \in A\}$$

[3] Si definisce spazio probabilizzabile una coppia $(\Omega, \mathcal{A})$, dove Ω è un insieme e $\mathcal{A}$ è una σ-algebra di eventi di Ω.

dove A è un generico elemento di $\mathcal{B}(\mathbb{R})$. Definiamo anche:
la **funzione di distribuzione congiunta** di ξ

$$F_\xi(x) = P\{\omega : \xi_1 \leq x_1, ..., \xi_n \leq x_n\} \quad x = (x_1, ..., x_n)$$

e le **funzioni di distribuzione marginali**

$$F_{\xi_k}(x_k) = P\{\omega : \xi_k \leq x_k\} \quad x_k \in \mathbb{R}, \ k = 1, ..., n.$$

Se una variabile aleatoria ξ può essere rappresentata come

$$\xi(\omega) = \sum_{i=1}^{+\infty} x_i I_{A_i}(\omega)$$

dove $\sum A_i = \Omega$, $A_i \in \mathcal{F}$ e $I_A(\omega)$ è l'indicatore di un insieme A di $\mathcal{F}$, si dice che è una **variabile aleatoria discreta**. Se la somma sopra è invece finita allora ξ è detta **variabile aleatoria (discreta) semplice**.
Si dice che ξ è una **variabile aleatoria continua** se la sua funzione di distribuzione $F_\xi(x)$ è continua per $x \in \mathbb{R}$.
Si dice che $\xi = (\xi_k)_{k=1,...,n}$ è una **variabile aleatoria vettoriale assolutamente continua** se c'è una funzione non negativa $f = f_\xi(x)$, chiamata densità tale che

$$F_\xi(x) = P\{\omega : \xi_1 \leq x_1, ..., \xi_n \leq x_n\} =$$
$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_\xi(y_1, y_2, ..., y_n) dy_1 dy_2 \cdots dy_n, \qquad x \in \mathbb{R}^n.$$

Questa definizione dipende da quale nozione di integrale si usa. Assumiamo l'integrale nel senso di Lebesgue.

Sia dato uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e un vettore aleatorio $X = (\xi_k)_{k=1,...,n}$ con $\xi_k : (\Omega, \mathcal{F}) \longmapsto (E_k, \mathcal{E}_k)$. Le n componenti di X sono **variabili aleatorie indipendenti** se vale:

$$P\{\xi_1 \in B_1, ..., \xi_n \in B_n\} = P\{\xi_1 \in B_1\} \dots P\{\xi_n \in B_n\}$$

comunque scelti $B_1 \in \mathcal{E}_1, ..., B_n \in \mathcal{E}_n$.

Infine diciamo che una variabile aleatoria $\xi : \Omega \longmapsto \mathbb{R}$ è una **variabile aleatoria integrabile** se

$$\int_\Omega |\xi| dP < +\infty \ .$$

Allora

$$E[\xi] = \int_\Omega \xi dP$$

esiste ed è detto **valore atteso** di ξ. Le proprietà del valore atteso che valgono nel caso discreto valgono anche nel caso generale.

4.5 Le variabili aleatorie assolutamente continue

Sia X una variabile aleatoria assolutamente continua. Per la definizione, risulta che

$$P\{X \in B\} = \int_B f_X(x)dx, \qquad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

Consideriamo il caso $n = 1$ e prendiamo il boreliano della forma $(a, b]$. Allora

$$P\{a < X \leq b\} = P\{X \leq b\} - P\{X \leq a\} = F_X(b) - F_X(a) = \int_a^b f_X(x)dx.$$

Inoltre abbiamo

$$P\{X \in (-\infty, +\infty)\} = \int_{-\infty}^{+\infty} f_X(x)dx = F_X(+\infty) - F_X(-\infty) =$$

$$= \lim_{x \to +\infty} F_X(x) - \lim_{x \to -\infty} F_X(x) = 1 - 0 = 1.$$

Distribuzione congiunta di variabili La densità congiunta di due variabili X, Y è la funzione $f_{X,Y}(x, y)$. Ci serve per poter determinare la seguente probabilità:

$$P\{X \in A, Y \in B\} = \int_B \int_A f_{X,Y}(x, y)dxdy.$$

Dalla conoscenza di $f_{X,Y}(x, y)$ possiamo determinare le densità marginali $f_X(x)$ e $f_Y(y)$:

$$P\{X \in A\} = P\{X \in A, Y \in (-\infty, +\infty)\} = \int_{-\infty}^{+\infty} \int_A f_{X,Y}(x, y)dxdy$$

inoltre

$$P\{X \in A\} = \int_A f_X(x)dx.$$

Concludiamo che

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy$$

e in modo analogo

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dx.$$

Se X e Y sono variabili indipendenti allora $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

Vediamo di seguito come si calcolano il valore atteso, i momenti di ordine k, la varianza e la covarianza di una variabile aleatoria assolutamente continua. Le proprietà di queste quantità descritte per il caso discreto valgono anche nel caso assolutamente continuo.

Valore atteso La variabile scalare X con densità $f_X(x)$ ammette valore atteso se e solo se $|x|f_X(x)$ è integrabile su $\mathbb{R}$ e in questo caso è definito da:

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

Per determinare il valore atteso di una variabile $g(X)$ (g è una funzione a valori reali) conoscendo $f_X(x)$, usiamo la seguente formula [4]:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx.$$

Momento di ordine k Il momento k-esimo o momento di ordine k di una variabile scalare X è la quantità, quando esiste finita

$$E[X^k] = \int_{-\infty}^{+\infty} x^k f_X(x) dx.$$

Varianza La varianza di una variabile scalare X che ammette momento secondo è la quantità

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

Covarianza La covarianza di due variabili X e Y è la quantità definita da:

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

4.5.1 Esempi di variabili assolutamente continue

Variabile aleatoria uniforme Si dice che X è una variabile aleatoria uniforme sull'intervallo $[a, b]$ (ma anche su (a, b), $[a, b)$ o $(a, b]$) se la sua densità è data da

$$f_X(x) = \begin{cases} 1/(b-a) & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases}$$

Indichiamo una tale X con $X \sim U(a, b)$. Più in generale diremo che X è una variabile aleatoria uniforme su $C \in \mathcal{B}(\mathbb{R}^d)$ (e scriviamo $X \sim U(C)$) se

$$f_X(x) = \frac{1}{m(C)} I_C(x)$$

[4] vedi Z. Brzeźniak-T. Zastawniak, **Basic Stochastic Processes**, Springer-Verlag, Londra, 1999, pagg.7,13.

dove per $m(C)$ intendiamo $m(C) = \int_C dx_1 \cdots dx_d$ Calcoliamo il valore atteso e varianza per il primo caso:

$$E[X] = \int_{-\infty}^{+\infty} x \frac{1}{b-a} I_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2}$$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 \frac{1}{b-a} I_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{a^2+ab+b^2}{3}$$

$$Var(X) = E[X^2] - (E[X])^2 = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{(b-a)^2}{12} .$$

Variabile aleatoria esponenziale Si dice che X è una variabile aleatoria esponenziale di parametro λ (e si scrive $X \sim Exp(\lambda)$) se la sua densità è

$$f_X(x) = \lambda e^{-\lambda x} I_{[0,+\infty)]}(x).$$

Calcoliamo il valore atteso:

$$E[X] = \int_{-\infty}^{+\infty} x\lambda e^{-\lambda x} I_{[0,+\infty)]}(x) dx = \int_0^{+\infty} x\lambda e^{-\lambda x} dx.$$

Risolviamo per parti l'integrale

$$\int x\lambda e^{-\lambda x} dx = -e^{-\lambda x} x + \int e^{-\lambda x} dx = -e^{-\lambda x} x - \frac{e^{-\lambda x}}{\lambda}$$

$$E[X] = \left[-e^{-\lambda x} x - e^{-\lambda x} \lambda^{-1}\right]_0^{+\infty} = \frac{1}{\lambda}.$$

Calcolo il momento secondo:

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 \lambda e^{-\lambda x} I_{[0,+\infty)]}(x) dx = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx.$$

Risolviamo per parti il seguente integrale:

$$\int x^2 \lambda e^{-\lambda x} dx = -x^2 e^{-\lambda x} - 2\frac{e^{-\lambda x}}{\lambda} x - 2\frac{e^{-\lambda}}{\lambda^2}$$

$$E[X^2] = \left[-x^2 e^{-\lambda x} - 2x e^{-\lambda x} \lambda^{-1} - 2e^{-\lambda x} \lambda^{-2}\right]_0^{+\infty} = \frac{2}{\lambda^2}.$$

Infine la varianza:

$$Var(X) = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Variabile aleatoria gamma Si dice che X è una variabile aleatoria gamma di parametri α e λ (e si scrive $X \sim \Gamma(\alpha, \lambda)$) se la sua densità è

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} I_{[0,+\infty)}(x)$$

e $\Gamma(\alpha)$ è: $\Gamma(\alpha) = \int_0^{+\infty} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx = (\alpha - 1)\Gamma(\alpha - 1)$. Determiniamo il valore atteso

$$E[X] = \int_0^{+\infty} x \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx = \frac{1}{\lambda \Gamma(\alpha)} \int_0^{+\infty} \lambda^{\alpha+1} x^\alpha e^{-\lambda x} dx =$$

$$= \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha\Gamma(\alpha)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda}$$

il momento secondo

$$E[X^2] = \int_0^{+\infty} x^2 \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx = \frac{1}{\lambda^2 \Gamma(\alpha)} \int_0^{+\infty} \lambda^{\alpha+2} x^{\alpha+1} e^{-\lambda x} dx =$$

$$= \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} = \frac{\alpha(\alpha+1)\Gamma(\alpha)}{\lambda^2\Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{\lambda^2}$$

e la varianza

$$Var(X) = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.$$

Ricordiamo che $\Gamma(1, \lambda) = Exp(\lambda)$.

Variabile aleatoria normale o gaussiana Si dice che X è una variabile aleatoria normale di parametri μ e σ^2 (e si scrive $X \sim N(\mu, \sigma^2)$) se la sua densità è

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

Determiniamo il valore atteso

$$E[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{\mu}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Consideriamo separatamente i due integrali. Nel primo poniamo $y = x - \mu$ e diventa

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} y e^{-y^2/2\sigma^2} dy = 0$$

per la simmetria. Il secondo invece si riscrive

$$\mu \int_{-\infty}^{+\infty} f(x)dx = \mu$$

per quanto visto in precedenza. Dunque

$$E[X] = \mu.$$

Determiniamo il momento secondo

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (y+\mu)^2 e^{-\frac{y^2}{2\sigma^2}} dy =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2\sigma^2}} dy + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} 2\mu y e^{-\frac{y^2}{2\sigma^2}} dy +$$

$$+ \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \mu^2 e^{-\frac{y^2}{2\sigma^2}} dy = \sigma^2 + 0 + \mu^2 = \sigma^2 + \mu^2.$$

Per calcolare il primo e per il terzo abbiamo usato il fatto che $\Gamma(1/2) = \sqrt{\pi}$, per il secondo la simmetria.

Infine la varianza

$$Var(X) = E[X^2] - (E[X])^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

Capitolo 5

Valore atteso condizionato: caso generale

Nel terzo capitolo abbiamo trattato il valore atteso condizionato nel caso discreto. In questo capitolo invece definiremo in generale il concetto di valore atteso condizionato rispetto ad una σ-algebra e vedremo come il caso discreto ne sia un caso particolare.

5.1 Condizionamento rispetto ad un evento

Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità, sia X una variabile aleatoria integrabile, definita su Ω, e B un evento di $\mathcal{F}$ tale che $P(B) \neq 0$. Il **valore atteso condizionato di X rispetto all'evento B** è definito da

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

Se $X = I_A$, funzione indicatrice dell'evento A ($A \in \mathcal{F}$), risulta $E[I_A|B] = P(A|B)$. Infatti dalla definizione

$$E[I_A|B] = \frac{1}{P(B)} \int_B I_A dP = \frac{1}{P(B)} \int_{B \cap A} dP = \frac{P(A \cap B)}{P(B)} = P(A|B)$$

5.2 Condizionamento rispetto ad una σ-algebra

Sia X una variabile integrabile definita su uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e sia $\mathcal{G}$ una σ-algebra tale che $\mathcal{G} \subseteq \mathcal{F}$. Il **valore atteso condizionato di X rispetto alla σ-algebra $\mathcal{G}$** è la variabile aleatoria $E[X|\mathcal{G}\,]$ tale che:

1. $E[X|\mathcal{G}\,]$ è $\mathcal{G}$-misurabile

2. per ogni $A \in \mathcal{G}$ si ha

$$\int_A X dP = \int_A E[X|\mathcal{G}\,] dP$$

La probabilità condizionata di un evento $A \in \mathcal{F}$ rispetto alla σ-algebra $\mathcal{G}$ può essere definita nel seguente modo:

$$P(A|\mathcal{G}) = E[I_A|\mathcal{G}]$$

La nozione di valore atteso condizionato rispetto a una σ-algebra estende quello di valore atteso rispetto ad una variabile aleatoria Y nel seguente modo:

$$E[X|\sigma(Y)] = E[X|Y]$$

dove $\sigma(Y)$ è la σ-algebra generata da Y.
Per definizione se $Y : (\Omega, \mathcal{F}) \longmapsto (E, \mathcal{E})$, allora $\sigma(Y)$ consiste di tutti gli insiemi della forma $\{Y \in B\}$, dove $B \in \mathcal{E}$.

Ci chiediamo se $E[X|\mathcal{G}\,]$ esiste ed è unico:

Proposizione 5.2.1. $E[X|\mathcal{G}\,]$ *esiste ed è unico nel senso che se* $X = X'$ *quasi sicuramente (cioè* $P(X = X') = 1$*) allora* $E[X|\mathcal{G}\,]$=$E[X'|\mathcal{G}\,]$ *quasi sicuramente.*

L'esistenza è garantita dal seguente teorema [1]:

Proposizione 5.2.2. (Teorema di Radon-Nikodym) *Sia* $(\Omega, \mathcal{F}, P)$ *uno spazio di probabilità e sia* $\mathcal{G}$ *una* σ*-algebra contenuta in* $\mathcal{F}$*. Allora per ogni variabile aleatoria integrabile* X *definita in* $(\Omega, \mathcal{F}, P)$ *esiste una variabile aleatoria* $\mathcal{G}$*-misurabile* Y *tale che*

$$\int_A X dP = \int_A Y dP$$

per ogni $A \in \mathcal{G}$.

L'unicità è conseguenza del seguente lemma [2]:

Lemma 5.2.1. *Sia* $(\Omega, \mathcal{F}, P)$ *uno spazio di probabilità e sia* $\mathcal{G}$ *una* σ*-algebra contenuta in* $\mathcal{F}$*. Se la variabile aleatoria* X *soddisfa alle seguenti condizioni:*

1. X *è* $\mathcal{G}$*-misurabile*

[1] Vedi P. Bilingsley, **Probability and Measure, 3th edition**, Wiley-Interscience Publication, USA, 1995, pag. 422.

[2] Vedi Z. Brzeźniak-T. Zastawniak, **Basic Stochastic Processes**, Springer-Verlag, Londra, 1999, pag.23.

2. *per ogni* $B \in \mathcal{G}$ $\int_B X dP = 0$

allora $X = 0$ *quasi sicuramente (cioè* $P\{X = 0\} = 1$*).*

Mostriamo che se X e X' sono variabili aleatorie definite in $(\Omega, \mathcal{F}, P)$ e tali che $X = X'$, allora $E[X|\mathcal{G}\,] = E[X'|\mathcal{G}\,]$. Infatti per definizione $E[X|\mathcal{G}\,]$ e $E[X'|\mathcal{G}\,]$ sono $\mathcal{G}$-misurabili, quindi anche $E[X|\mathcal{G}\,] - E[X'|\mathcal{G}\,]$ è variabile aleatoria $\mathcal{G}$-misurabile. Vediamo se la seconda condizione del lemma è soddisfatta: per ogni $B \in \mathcal{G}$

$$\int_B (E[X|\mathcal{G}\,] - E[X'|\mathcal{G}\,])dP = \int_B E[X|\mathcal{G}\,]dP - \int_B E[X'|\mathcal{G}\,]dP =$$

$$= \int_B X dP - \int_B X' dP = \int_B (X - X')dP = 0$$

Dunque per il lemma $E[X|\mathcal{G}\,] = E[X'|\mathcal{G}\,]$. Vediamo le proprietà del valore atteso condizionato.

Proposizione 5.2.3. *Siano* $a, b \in \mathbb{R}$, X *e* Y *variabile aleatorie integrabili definite in uno spazio di probabilità* $(\Omega, \mathcal{F}, P)$. *Siano inoltre* $\mathcal{G}$ *e* $\mathcal{H}$ σ*-algebre contenute in* $\mathcal{F}$. *Allora:*

1. $E[aX + bY|\mathcal{G}\,] = aE[X|\mathcal{G}\,] + bE[Y|\mathcal{G}\,]$*:*

2. *se* X *è* $\mathcal{G}$*-misurabile e* XY *è integrabile (*$\int_\Omega |XY|dP < +\infty$*) allora* $E[XY|\mathcal{G}\,] = XE[Y|\mathcal{G}\,]$*;*

3. *se* X *e* $\mathcal{G}$ *sono indipendenti (cioè* $\sigma(X)$ *e* $\mathcal{G}$ *- entrambe contenute in* $\mathcal{F}$ *- sono indipendenti: per ogni* $A \in \sigma(X)$ *e* $B \in \mathcal{G}$ *sono indipendenti) allora* $E[X|\mathcal{G}\,] = E[X]$*;*

4. *se* $\mathcal{H} \subset \mathcal{G}$ *allora* $E[E[X|\mathcal{G}\,]|\mathcal{H}\,] = E[X|\mathcal{H}\,]$*;*

5. $E[E[X|\mathcal{G}]] = E[X]$*;*

6. *se* $X \geq 0$ *quasi sicuramente allora* $E[X|\mathcal{G}\,] \geq 0$ *quasi sicuramente.*

Dimostrazione.

1. Per ogni $B \in \mathcal{G}$ abbiamo per definizione

$$\int_B E[aX + bY|\mathcal{G}\,]dP = \int_B (aX + bY)dP$$

e

$$\int_B (aE[X|\mathcal{G}\,] + bE[Y|\mathcal{G}\,])dP = a\int_B E[X|\mathcal{G}\,]dP + b\int_B E[Y|\mathcal{G}\,]dP =$$

$$= a\int_B XdP + b\int_B YdP = \int_B (aX+bY)dP$$

Per l'unicità del valore atteso condizionato si ha l'uguaglianza

$$E[aX+bY|\mathcal{G}\,] = aE[X|\mathcal{G}\,] + bE[Y|\mathcal{G}\,].$$

2. Prima verifichiamo per il caso $X = I_A$, $A \in \mathcal{G}$. Per ogni $B \in \mathcal{G}$

$$\int_B I_A E[Y|\mathcal{G}\,]dP = \int_{A\cap B} E[Y|\mathcal{G}\,]dP = \int_{A\cap B} YdP = \int_B I_A Y dP$$

Per la definizione di valore atteso condizionato e per la sua unicità otteniamo

$$I_A E[Y|\mathcal{G}\,] = E[I_A Y|\mathcal{G}\,]$$

Consideriamo ora il caso in cui X sia una funzione a scalino $\mathcal{G}$-misurabile. Allora possiamo scrivere per definizione

$$X = \sum_{k=1}^{m} a_k I_{A_k}$$

dove $A_k \in \mathcal{G}$ per $k = 1, 2, ..., m$. Basta ora applicare la prima proprietà e si ottiene l'uguaglianza. Infine per il caso più generale occorre approssimare X ad una funzione a scalino $\mathcal{G}$-misurabile.

3. Osserviamo che per ogni $B \in \mathcal{G}$ le variabili X e I_B sono indipendenti (cioè per ogni $F_1, F_2 \in \mathcal{F}$ gli eventi $\{X \in F_1\}$ e $\{I_B \in F_2\}$ sono indipendenti) essendo X indipendente da $\mathcal{G}$. Per la proprietà 5 della proposizione 2.5.2, che vale anche nel caso generale, abbiamo $E[XI_B] = E[X]E[I_B]$. Dunque

$$\int_B E[X]dP = E[X]E[I_B] = E[XI_B] = \int_B XdP$$

e per l'unicità vale l'uguaglianza $E[X|\mathcal{G}\,] = E[X]$.

4. Dalla definizione abbiamo che per ogni $B \in \mathcal{G}$

$$\int_B E[X|\mathcal{G}\,]dP = \int_B XdP$$

e anche per $B \in \mathcal{H}$

$$\int_B E[E[X|\mathcal{G}\,]|\mathcal{H}\,]dP = \int_B E[X|\mathcal{G}\,]dP = \int_B XdP$$

e naturalmente
$$\int_B E[X|\mathcal{H}\,]dP = \int_B XdP$$
Poiché $\mathcal{H} \subset \mathcal{G}$, per ogni $B \in \mathcal{H}$ si ha
$$\int_B E[E[X|\mathcal{G}\,]|\mathcal{H}\,]dP = \int_B E[X|\mathcal{G}\,]dP = \int_B E[X|\mathcal{H}\,]dP$$
Dunque otteniamo
$$E[E[X|\mathcal{G}\,]|\mathcal{H}\,] = E[X|\mathcal{H}\,].$$

5. È un caso speciale della quarta proprietà quando $\mathcal{H} = \{\varnothing, \Omega\}$. Mostriamo dapprima che $E[X|\mathcal{H}] = E[X]$. Per definizione di valore atteso abbiamo
$$E[X] = \int_\Omega XdP$$
e inoltre essendo $E[X]$ una variabile aleatoria costante:
$$E[X] = E[X] \cdot 1 = E[X]P(\Omega) = \int_\Omega E[X]dP$$
Per l'altro evento di $\mathcal{H}$ abbiamo
$$\int_\varnothing XdP = 0 = \int_\varnothing E[X]dP$$
In conclusione $E[X|\mathcal{H}\,] = E[X]$. Allo stesso modo vale $E[E[X|\mathcal{G}]] = E[X]$ perché per una tale $\mathcal{H}$ $E[E[X|\mathcal{G}\,]|\mathcal{H}\,] = E[E[X|\mathcal{G}]]$ e per la quarta proprietà $E[E[X|\mathcal{G}\,]|\mathcal{H}\,] = E[X|\mathcal{H}\,] = E[X]$.

6. Per ogni $n \in \mathbb{N}$ poniamo
$$A_n = \{E[X|\mathcal{G}\,] \leq -\frac{1}{n}\}$$
Allora $A_n \in \mathcal{G}$. Possiamo scrivere
$$\{E[X|\mathcal{G}\,] \leq 0\} = \bigcup_{n=1}^{+\infty} A_n.$$
Se $X \geq 0$ quasi sicuramente, segue
$$0 \leq \int_{A_n} XdP = \int_{A_n} E[X|\mathcal{G}\,]dP \leq -\int_{A_n} \frac{1}{n}dP = -\frac{1}{n}P(A_n).$$

Significa che $P(A_n) = 0$. Infine

$$P\{E[X|\mathcal{G}\,] \leq 0\} = P(\bigcup_{n=1}^{+\infty} A_n) = \lim_{n\to+\infty} P(A_n) = 0$$

Abbiamo usato la seguente proprietà [3]:

Data una successione $(A_n)_{\mathbb{N}}$, $A_n \in \mathcal{G}$ per ogni $n \in \mathbb{N}$, crescente e tale che $\bigcup_n A_n \in \mathcal{G}$, risulta

$$\lim_{n\to+\infty} P(A_n) = P(\bigcup_{n=1}^{+\infty} A_n)$$

Dunque $P\{E[X|\mathcal{G}\,] \geq 0\} = 1$.

5.3 Condizionamento rispetto ad una variabile aleatoria

Siano X e Y variabile aleatorie definite su uno stesso spazio di probabilità $(\Omega, \mathcal{F}, P)$. Il **valore atteso condizionato di X rispetto a Y** è la variabile $E[X|Y]$ tale che

1. $E[X|Y]$ è σ(Y)-misurabile
2. per qualsiasi evento $A \in \sigma(Y)$

$$\int_A E[X|Y]dP = \int_A XdP$$

Possiamo anche definire la probabilità condizionata di un evento $A \in \mathcal{F}$ rispetto alla variabile Y:

$$P(A|Y = y) = E[I_A|Y = y].$$

5.3.1 Esempio di applicazione

Prendiamo come spazio di probabilità $(\Omega, \mathcal{F}, P)$ il seguente: $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$ e P la misura di Lebesgue su $[0, 1]$. Vogliamo determinare $E[\xi|\eta]$ per

$$\xi(x) = 2x^2 \qquad \eta(x) = \begin{cases} 2 & \text{se } x \in [0, \frac{1}{2}) \\ x & \text{se } x \in [\frac{1}{2}, 1] \end{cases}.$$

[3] Vedi **N. Cufaro Petroni**, *Lezioni di Calcolo delle Probabilità*, Edizioni dal Sud, Modugno(Bari), 1996, pag.130.

Descriviamo $\sigma(\eta)$. Per ogni boreliano $B \subset [\frac{1}{2}, 1]$ abbiamo

$$B_1 = \{\eta \in B\} \in \sigma(\eta)$$

e

$$B_2 = [0, \frac{1}{2}) \cup B_1 = \{\eta = 2\} \cup \{\eta \in B\} \in \sigma(\eta).$$

Gli insiemi di questi due tipi esauriscono tutti gli elementi di $\sigma(\eta)$.

Se $E[\xi|\eta]$ è $\sigma(\eta)$-misurabile, deve essere costante in $[0, \frac{1}{2}]$ perché η lo è. Possiamo scrivere la seguente corrispondenza di eventi: $\{\eta = 2\} = [0, \frac{1}{2})$. Se per ogni $x \in [0, \frac{1}{2})$

$$E[\xi|\eta](x) = E[\xi|[0, \frac{1}{2})] = \frac{1}{P([0, \frac{1}{2}))} \int_{[0,\frac{1}{2})} \xi dP = \frac{1}{P([0, \frac{1}{2}))} \int_{[0,\frac{1}{2})} \xi(x)dx =$$

$$= \frac{1}{\frac{1}{2}} \int_0^{\frac{1}{2}} 2x^2 dx = 2\frac{1}{12} = \frac{1}{6}$$

allora è soddisfatta la seguente condizione del valore medio condizionato:

$$\int_{[0,\frac{1}{2})} E[\xi|\eta](x)dx = \int_{[0,\frac{1}{2})} \xi(x)dx = \frac{1}{12}.$$

Se $E[\xi|\eta] = \xi$ in $[\frac{1}{2}, 1]$ allora sicuramente per ogni $B \subset [\frac{1}{2}, 1]$ abbiamo

$$\int_B E[\xi|\eta](x)dx = \int_B \xi(x)dx.$$

Dunque abbiamo trovato che

$$E[\xi|\eta](x) = \begin{cases} \frac{1}{6} & \text{se } x \in [0, \frac{1}{2}) \\ 2x^2 & \text{se } x \in [\frac{1}{2}, 1] \end{cases}$$

Poiché ogni elemento di $\sigma(\eta)$ è della forma B_1 o B_2 allora le condizioni della definizione di valore atteso condizionato sono entrambe soddisfatte.

Bibliografia

- P. Bilingsley, *Probability and Measure, 3th edition*, Wiley-Interscience Publication, USA, 1995

- Z. Brzeźniak, T. Zastawniak, *Basic Stochastic Processes: A Course Through Exercises*, Springer-Verlag, London, 1999

- N. Cufaro Petroni, *Lezioni di Calcolo delle Probabilità*, Edizioni dal Sud, Modugno(Bari), 1996

- P. Dai Pra, *Dispense di Calcolo delle Probabilità*, anno accademico 2002-2003

- G. De Marco, *Analisi Due*, Decibel editrice, Padova, 1999

- G.R. Grimmett, D. Stirzaker, *Probability and Random Processes*, Clarendon Press, Oxford, 1982

- J. Hoffmann-Jørgensen, *Probability with a view toward statistics*, volume I, Chapman & Hall, USA, 1994

- S.M. Ross, *A First Course in Probability, Sixth Edition*, Prentice Hall, Upper Saddle River (USA), 2002

- S.M. Ross, *Introduction to Probability Models, Eighth Edition*, Academic Press, USA, 2003

- Y.A. Rozanov, *Probability Theory: a concise course*, Dover Publication, New York, 1977

- A.N. Shiryayev, *Probability*, Springer-Verlag, USA, 1984

- Y.G. Sinai, *Probability Theory: An Introductory Course*, Springer-Verlag, Berlin Heidelberg, 1992