4.4. Deviazioni dalla media

Il valore atteso $E(X)$ di una variabile aleatoria X dà un'idea di quanto ci possiamo aspettare, ma abbiamo già visto che non è esattamente il valore $E(X)$ che ci possiamo attendere come risultato di una realizzazione di X. Il teorema di De Moivre-Laplace fornisce una stima molto precisa di quanto i valori che si realizzano più spesso differiscono dal valore atteso di una distribuzione di Bernoulli, ma per il caso generale discutiamo ora una prima stima. Ricordiamo inoltre che non abbiamo ancora chiarito il significato del termine $\sqrt{p(1-p)}$.

Lo scarto massimo $\max(X - E(X))$ tende a sopravvalutare lo scarto tipico di una realizzazione e chiaramente lo scarto medio $E(X - E(X))$ non ha significato, essendo identicamente nullo per la linearità del valore atteso. Lo scarto assoluto medio $E(|X - E(X)|)$ è una soluzione migliore.

Esempio 40. *Per $X \sim B(1,p)$ si ha $E(X) = p$ e lo scarto assoluto medio risulta essere $2p(1-p)$; per il risultato del lancio di un dado Y lo scarto assoluto medio è $1,5$.*

Si osservi però che il calcolo dello scarto assoluto medio per variabili con distribuzione $B(n,p)$ risulta laborioso e che comunque non risulta spiegato il termine $\sqrt{p(1-p)}$. Pensandoci, non c'è una ragione evidente per non considerare lo scarto quadratico medio di X

$$SD(X) = \sqrt{E(X - E(X))}$$

detta anche **deviazione standard** di X. O anche qualche altra potenza. Tuttavia risulta che

Esempio 41. *Per $X \sim B(1,p)$ lo scarto quadratico medio risulta essere $\sqrt{p(1-p)}$ e per il risultato del lancio di un dado lo scarto quadratico medio è circa $1,70$.*

Si vede da questi esempi che lo scarto quadratico medio non differisce sostanzialmente dallo scarto assoluto medio, ma soprattutto 'spiega' il termine che appare nel teorema di De Moivre-Laplace. L'espressione $Var(X) = E(X - E(X))$ è detta **varianza** di X. La varianza può essere anche calcolata come segue.

Lemma 9. *Per ogni variabile aleatoria X si ha*

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

Dimostrazione.

$$\begin{aligned} Var(X) &= E(X - E(X))^2 \\ &= E(X^2 - 2XE(X) + (E(X))^2) \\ &= E(X^2) - (E(X))^2 \end{aligned} \tag{4.16}$$

□

Un importante risultato relativo alla varianza è che risulta additiva per somme di variabili aleatorie indipendenti.

TEOREMA 9. *Per ogni $X_1, \ldots X_n$ variabili aleatorie indipendenti su uno spazio di probabilità (S, P) si ha*

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

e

$$SD(\sum_{i=1}^{n} X_i) = \sqrt{\sum_{i=1}^{n} Var(X_i)}.$$

DIMOSTRAZIONE. Iniziamo da $n = 2$.

$$\begin{aligned}
Var(X_1 + X_2) &= E(X_1 + X_2)^2 - (E(X_1 + X_2))^2 \\
&= E(X_1)^2 + E(X_2)^2 + 2E(X_1X_2) \\
&\qquad -(E(X_1))^2 - (E(X_2))^2 - 2E(X_1)E(X_2) \\
&= Var(X_1) + Var(X_2) + 2(E(X_1X_2) - E(X_1)E(X_2)) \\
&= Var(X_1) + Var(X_2)
\end{aligned}$$

poiché per variabili indipendenti $E(X_1X_2) = E(X_1)E(X_2)$ dal Teorema 8.

Il risultato per n generico si ottiene per induzione essendo $\sum_{i=1}^{n-1} X_i$ indipendente da X_n.

Il risultato per la deviazione standard segue dalla definizione. □

Quindi per variabili i.i.d. $X_1, \ldots, X_n$ si ha che $SD(\sum_{i=1}^{n} X_i) = \sqrt{nSD(X_1)}$. In particolare se $X_i \sim B(1, p)$ allora $SD(\sum_{i=1}^{n} X_i) = \sqrt{np(1-p)}$.

Quindi la deviazione standard è contemporaneamente facile da calcolare per somme di variabili indipendenti ed assegna un significato all'espressione $\sqrt{np(1-p)}$ che appare nel teorema di De Moivre-Laplace. Questo ci permette di congetturare un'estensione di questo risultato a tutte le variabili aleatorie finite. Rileggendo infatti il risultato per

$T_n \sim B(n,p)$ ossia tale che $T_n = \sum_{i=1}^n X_i$, $X_i \sim B(1,p)$, si ha

$$\begin{aligned}
\sum_{j \geq np + a\sqrt{np(1-p)}} \mathcal{P}(j,n,p) &= \sum_{j \geq np + a\sqrt{np(1-p)}} P(\sum_{i=1}^n X_i = j) \\
&= P\left(\sum_{i=1}^n X_i \geq np + a\sqrt{np(1-p)}\right) \\
&= P\left(\sum_{i=1}^n X_i \geq nE(X_1) + a\sqrt{nVar(X_1)}\right) \\
&= P\left(\frac{\sum_{i=1}^n X_i - nE(X_1)}{\sqrt{nVar(X_1)}} \geq a\right) \\
&= P\left(\frac{(\sum_{i=1}^n X_i) - E(\sum_{i=1}^n X_i)}{\sqrt{Var(\sum_{i=1}^n X_i)}} \geq a\right).
\end{aligned}$$

L'espressione $\frac{(\sum_{i=1}^n X_i)-E(\sum_{i=1}^n X_i)}{\sqrt{Var(\sum_{i=1}^n X_i)}}$ può essere interpretata in un modo diretto: $\sum_{i=1}^n X_i$ è una variabile aleatoria che dà i possibili valori della media empirica, a questa viene sottratto il suo valore atteso dividendo poi per la sua deviazione standard.

Definizione 12. *Data una variabile aleatoria X, si dice* **variabile standardizzata** *o* **versione standardizzata di** X *l'espressione*

$$\frac{X - E(X)}{SD(X)}$$

.

Il Teorema di De Moivre-Laplace ci dice quindi che la versione standardizzata della somma di variabili aleatorie indipendenti (binomiali) ha probabilità descritte asintoticamente dalla gaussiana. La versione standardizzata della somma è ben definita per ogni variabile aleatoria finita, quindi è ragionevole congetturare che

Teorema 10 (Teorema Centrale del Limite). *Per variabili aleatorie i.i.d. $X_1, \ldots, X_n$ su uno spazio di probabilità finito vale per ogni $a \in \mathbb{R}$*

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_i - nE(X_1)}{\sqrt{nVar(X_1)}} \geq a\right) = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Per ora non dimostreremo questo teorema poiché ne vale in realtà una versione ancora più generale.

Ci si potrebbe porre il problema di stimare la deviazione della deviazione dal valor medio dalla sua media, per esempio con $E(|X - E(X) -$

$\sqrt{E(X-E(X))^2}|)$ oppure $E(\sqrt{(X-E(X)-\sqrt{E(X-E(X))^2})^2})$. Tuttavia questo non viene generalmente sviluppato perché (1), come illustrato dal Teorema centrale del limite, spesso è il secondo momento, e quindi la deviazione standard, che descrive in dettaglio la distribuzione della deviazione dalla media, (2) non introduce nessuna novità teorica perché si tratta sempre di un valore atteso di una deviazione e (3) diventa più chiaro porsi direttamente il problema della ricostruzione di una variabile aleatoria dalla conoscenza dei suoi momenti, un problema che non trattiamo in queste note.

4.5. Diseguaglianze e legge debole dei grandi numeri

La varianza dà un'idea della deviazione tipica dalla media, ma ora vediamo qualche risultato che dia una stima di questa deviazione.

LEMMA 10 (Diseguaglianza di Markov). *Per ogni variabile aleatoria $X \geq 0$ non negativa e per ogni $a > 0$ si ha*

$$P(X \geq a) \leq \frac{E(X)}{a}$$

.

DIMOSTRAZIONE. Essendo $X \geq 0$, per ogni $a > 0$ si ha:

$$\begin{aligned} E(X) &= \sum_{x \in S_X} x P_X(x) \\ &\geq \sum_{x \in S_X, x \geq a} x P_X(x) \geq a P(X \geq a) \end{aligned}$$

□

Da questo segue

COROLLARIO 2 (Diseguaglianza di Chebyshev). *Per ogni variabile aleatoria X finita e per ogni $a > 0$ si ha:*

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2},$$

ossia

$$P(\left|\frac{X-E(X)}{SD(X)}\right| \geq a) \leq \frac{1}{a^2}.$$

DIMOSTRAZIONE. Essendo $(|X-E(X)|)^2 = (X-E(X))^2 \geq 0$, dal Lemma 10 segue che per ogni $a > 0$ si ha:

$$P(|X-E(X)| \geq a) = P((X-E(X))^2 \geq a^2) \leq \frac{Var(X)}{a^2}$$

□

Queste diseguaglianze non sono troppo accurate, come si vede dall'esempio seguente, anzi talvolta sono banali in quanto il maggiorante è maggiore di 1.

Esempio 42. *Se X è il risultato del lancio di un dado allora*

$$\frac{1}{3} = P(X \geq 5) \leq \frac{1}{5}3,5 = 0,7$$

dal Lemma 10 e

$$\frac{1}{3} = P(|X - 3,5| \geq 2,5) \leq \frac{1}{(2,5)^2}Var(X) \approx 0.47$$

dal Corollario 2.

Si possono tuttavia dedurre due cose.

Lemma 11. *Se $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = 1$ allora la costante nell'approssimazione di Stirling soddisfa $e^c = \sqrt{2\pi}$.*

Dimostrazione. Dalla parte dimostrata della legge di De Moivre-Laplace scritta in termini di variabili aleatorie sappiamo che per ogni $a_1, a_2 \geq 0$

$$Q_n(a_1, a_2) = P(\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \in [a_1, a_2]) \to_{n\to\infty} \int_{a_1}^{a_2} \frac{1}{e^c} e^{-x^2/2}dx$$

e per la simmetria della distribuzione di Bernoulli e della gaussiana, questo risultato vale anche quando $a_1, a_2 \leq 0$ e dall'additività per ogni $a_1, a_2 \in \mathbb{R}$. A noi interessa qui per $a_1 = -a = -a_2 < 0$.
1. Assumendo $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = 1$ si ha che per ogni $a > 0$ vale $1 \geq Q_n(-a, a)$ per ogni n; pertanto anche

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = 1 &\geq \lim_{n\to\infty} Q_n(-a, a) \\ &= \int_{-a}^{a} \frac{1}{e^c} e^{-x^2/2}dx \end{aligned} \tag{4.17}$$

per ogni $a > 0$. Questo implica $e^c \geq \sqrt{2\pi}$.
2. D'altra parte, dalla definizione di limite, per ogni $\epsilon > 0$ esiste $N > 0$ tale che per ogni $n \geq N$ vale la prima diseguaglianza in

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx &= 1 = P(\sum_{i=1}^n X_i \in [0, n]) \\ &= P(\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \in [-a, a]) \\ &\qquad -P(|\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}}| \geq a) \\ &\leq \int_{-a}^{a} \frac{1}{e^c} e^{-x^2/2}dx + \epsilon + P(|\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}}| \geq a) \\ &\leq \int_{-\infty}^{\infty} \frac{1}{e^c} e^{-x^2/2}dx + \epsilon + \frac{1}{a^2}, \end{aligned}$$

in cui l'ultima diseguaglianza segue dalla diseguaglianza di Chebyshev. Quindi se $a > \epsilon^{-1/2}$ si ha

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \leq \int_{-\infty}^{\infty} \frac{1}{e^c} e^{-x^2/2} dx + 2\epsilon$$

da cui $e^c \leq \sqrt{2\pi}$. □

La seconda conseguenza riguarda la probabilità di una deviazione dal valore atteso dell'ordine di n:

TEOREMA 11 (Legge (debole) dei grandi numeri). *Per variabili aleatorie finite i.i.d.* $X_1, X_2, \ldots$ *su si ha che per ogni* $\alpha > 0$

$$\lim_{n\to\infty} P(\left|\sum_{i=1}^{n} X_i - E(X_1)\right| > \alpha) = 0$$

nel senso che fissata una distribuzione finita (S_X, P_X) *per ogni* $\epsilon > 0$ *esiste* N *tale che per ogni* $n \geq N$ *se si prendono* n *variabili aleatorie indipendenti ognuna con distribuzione* (S_X, P_X) *allora* $P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right| > \alpha) < \epsilon$.

L'esistenza di variabili indipendenti con una distribuzione data verrà verificata più avanti.

DIMOSTRAZIONE. Dalla diseguaglianza di Chebyshev si ha:

$$\begin{aligned} P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right| > \alpha) &= P(\left|\sum_{i=1}^{n} X_i - nE(X_1)\right| > n\alpha) \\ &\leq \frac{1}{n^2\alpha^2} Var(\sum_{i=1}^{n} X_i) \\ &= \frac{nVar(X_1)}{n^2\alpha^2} \to_{n\to\infty} 0 \end{aligned}$$

□

ESEMPIO 43. *Se* X_i *sono i risultati di lanci indipendenti di un dado allora*

$$P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right| > 10^{-10}) \to_{n\to\infty} 0$$

anzi

$$P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right| > 10^{-10}) \leq \frac{1,70 \times 10^{-10}}{n}.$$

Si noti che quest'ultima stima ha senso solo per $n \geq 1,70 \times 10^{10}$*; tuttavia ha poco senso dare troppa importanza al valore quantitativo di queste stime a causa della loro scarsa accuratezza.*

4.6. Approssimazione di Poisson

Vediamo ora un'approssimazione per la distribuzione binomiale quando la probabilità di successo p ed il numero di prove n sono tali che pn è dell'ordine di 1:

Esempio 44. *Sorteggiando con reinserimento dalla tombola* 180 *volte la probabilità che l'*1 *esca esattamente* 2 *volte è :*

$$\mathcal{P}(2, 180, \frac{1}{90}) = \binom{180}{2} \frac{1}{90^2} (\frac{89}{90})^{178}.$$

Teorema 12 (Approssimazione di Poisson). *Se* $p = p_n$ *è tale che* $\lim_{n\to\infty} np_n = \lambda > 0$ *si ha*

$$\lim_{n\to\infty} \mathcal{P}(k, n, p_n) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Dimostrazione. poiché $\lim_{n\to\infty}(1 - p_n)^n = e^{-\lambda}$ si ha

$$\begin{aligned} \mathcal{P}(k, n, p_n) &= \binom{n}{k} (p_n)^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1)\ldots(n-k+1)}{k!} (p_n^k)(1 - p_n)^n (1 - p_n)^{-k} \\ &\to_{n\to\infty} \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

□

CAPITOLO 5

PROBABILITÀ SU INSIEMI DISCRETI

5.1. Spazi di probabilità su insiemi discreti

Contemporaneamente allo sviluppo delle probabilità finite alcuni studiosi si resero conto che taluni problemi non si potevano formalizzare con un numero finito di possibilità. I primi problemi di questo tipo appaiono in un libro di Huygens del 1657.

Esempio 45. *Se giocando a dadi A vince se esce prima il* 6 *di* 1 *e* 2*, e viceversa per B, qual è la probabilità che A vinca? Chiaramente è naturale considerare la probabilità che vinca A al lancio k, per $k \in \mathbb{N}$.*

Ciò motiva l'introduzione di spazi di probabilità con S numerabile; ma questo introduce una nuova scelta: se $\{A_i\}_{i\in\mathbb{N}}$ è una famiglia numerabile di eventi disgiunti, si dovrà richiedere la numerabile additività della probabilità o solo quella finita? La questione non è risolta in modo univoco: una condizione più stringente limita il campo di applicazione della teoria ma ne semplifica gli sviluppi ed è quindi ragionevole richiederla quando il campo di applicazione resti comunque sufficientemente ampio. In generale, l'imposizione dell'additività numerabile non pone restrizioni di rilievo alla applicazioni fisiche (trattandosi perlopiù di esperimenti ripetibili a piacere) mentre ne pone in ambiti economici, trattandosi spesso in quel caso di situazioni solo occasionalmente ripetibili. Noi qui lo adotteremo sia per semplicità sia perché comunque la teoria qui esposta ha una sufficiente ampiezza di applicazioni anche in campo socio-economico-finanziario.

Definizione 13. *Uno* **spazio di probabilità discreto** *è una coppia (S, P) in cui S è un insieme al più numerabile e P è una funzione definita sulle parti di S tale che:*
(1) $P(S) = 1$
(2) per ogni $A \subseteq S$, $P(A) \in [0, 1]$;
(3) se A_i, $i = 1, 2, \dots$, $A_i \subseteq S$ sono insiemi disgiunti allora

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

in cui si intende che a destra c'è una serie a termini positivi convergente.

La richiesta di additività numerabile rende in realtà il modello più semplice di quanto l'assiomatica appena descritta sembrasse prospettare.

LEMMA 12. *Tutti e soli gli spazi di probabilità discreti sono ottenuti da un insieme al più numerabile S e da una funzione $q : S \to \mathbb{R}^+$ tale che $\sum_{s\in S} q(s) < \infty$ ponendo per ogni $A \subseteq S$*

$$P(A) = \frac{\sum_{s\in A} q(s)}{\sum_{s\in S} q(s)}.$$

DIMOSTRAZIONE. Ogni spazio di probabilità discreto (S, P) si può rappresentare come detto con $q(s) = P(s)$. Viceversa, dato q come nell'asserzione basta porre $P(s) = \frac{q(s)}{\sum_{s\in S} q(s)}$; la verifica che P è una probabilità è lasciata per esercizio. $\square$

OSSERVAZIONE 7. *Per gli spazi di probabilità discreti vale l'additività finita della probabilità e quindi valgono tutti i risultati dei capitoli 2 e 3 nella forma in cui sono enunciati, ossia riferiti ad un numero finito di eventi. E' solo quando è coinvolta una famiglia numerabile di eventi che dobbiamo dedurre i risultati dall'additività numerabile (e dalle proprietà delle serie).*

ESEMPIO 46. *Nell'esempio 45 se A è l'evento che vince A; A_i è l'evento che vince A alla i-sima prova; N_j è l'evento che alla j-sima prova non escono nessuno di $\{1, 2, 6\}$ e S_j è l'evento che esce il 6 alla j-sima prova per l'indipendenza della prove si ha*

$$P(A_i) = P(\cap_{j=1}^{i-1} N_j \cap S_i) = (\frac{1}{2})^{i-1}\frac{1}{6}$$

e quindi

$$\begin{aligned} P(A) &= P(\cup_{i=1}^{\infty} A_i) \\ &= \sum_{i=1}^{\infty} P(A_i) \\ &= \sum_{i=1}^{\infty} (\frac{1}{2})^{i-1}\frac{1}{6} = \frac{1}{3}. \end{aligned}$$

La probabilità che il primo successo di prove indipendenti ognuna con probabilità di successo p avvenga alla n-sima prova si può calcolare considerando, come nell'esempio, l'evento A_i che il primo successo sia alla i-sima prova e l'evento B_j che indica successo alla j-sima prova; per l'indipendenza della prove si ha

$$P(A_i) = P(\cap_{j=1}^{i-1} B_j^c \cap B_i) = (1-p)^{i-1}p.$$

Ora questo suggerisce di definire uno spazio di probabilità discrete $(\overline{S}, \overline{P})$ con $\overline{S} = \mathbb{N}$ e $\overline{P}$ data da

$$\overline{P}(i) = (1-p)^{i-1}p.$$

Infatti correttamente si ha;

$$\sum_{i=1}^{\infty} \overline{P}(i) = \sum_{i=1}^{\infty} (1-p)^{i-1} p = 1$$

.

DEFINIZIONE 14. *Lo spazio di probabilità $(\overline{S}, \overline{P})$ relativo al primo successo in prove indipendenti ognuna con probabilità di successo p è detto* **distribuzione geometrica** *di parametro p.*

Si noti che anche l'approssimazione di Poisson generava una funzione di k per $k \in \mathbb{N}$; inoltre poiché

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1$$

si possono prendere questi valori come probabilità:

DEFINIZIONE 15. *Per ogni $\lambda \in \mathbb{R}$, uno spazio di probabilità $(\overline{S}, \overline{P})$ con $\overline{S} = \mathbb{N}$ e $\overline{P}$ data da*

$$\overline{P}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

è detto **distribuzione di Poisson** *di parametro λ.*

5.2. Variabili aleatorie discrete

Passiamo ora allo studio delle variabili aleatorie definite su uno spazio di probabilità discreto. Non c'è nessuna difficoltà a porre

DEFINIZIONE 16. *Dato uno spazio di probabilità (S, P), una variabile aleatoria discreta X è una funzione $X : S \to \mathbb{R}$.*

Anche per una variabile aleatoria discreta X è possibile definire la distribuzione (S_X, P_X) come in (4.12).

ESERCIZIO 13. *Verificare che (S_X, P_X) è uno spazio di probabilità discreto.*

Si può poi ripetere la Definizione 6 di uguaglianza in distribuzione.

ESEMPIO 47. *Se (S, P), $S = \mathbb{N}$ è uno spazio di probabilità che descrive la distribuzione del primo successo in prove indipendenti con probabilità di successo p, la variabile aleatoria Y = tempo del primo successo è definita da $X(k) = k$ ha distribuzione $P(X = j) = (1 - p)^{j-i} p$ é detta variabile geometrica(p), mentre la variabile aleatoria Y = tempo di attesa del primo successo è definita da $Y(k) = k - 1$ ha distribuzione $P(Y = j) = (1-p)^j p$.*

Si noti che $Y =^d X - 1$.

ESEMPIO 48. *Se (S_λ, P_λ) è uno spazio di probabilità di Poisson and una variabile aleatoria N ha distribuzione (S_λ, P_λ) allora si dice che N ha distribuzione di Poisson(λ).*

C'è invece qualche problema nel definire il valore atteso:

Esempio 49. *In un gioco, se la prima uscita del* 6 *in un dado è alla* k*-sima prova si vince (o si perde se negativo) l'importo* x_k. *Giochereste se*

(a) $x_k = (-1)^k k$*?*

(b) $x_k = (-1)^k (6/5)^k$*?*

Si consideri uno spazio di probabilità geometrico(p) e la variabile aleatoria $X(k) = (x_k)$. *Per valutare il nostro vantaggio nel gioco verrebbe di calcolare* $E(X)$*, ma una ragionevole espressione sarebbe* $\sum_{k=1}^{\infty} x_k (5/6)^k$ *che nel caso (a) dà* $11/375 - 6/11 < 0$*, ma nel caso (b) dà* $1/5 \sum_{k=1}^{\infty} (-1)_k$ *che è una serie indeterminata.*

Per non rischiare di incontrare situazioni come questa e per garantire che le principali proprietà del valore atteso siano conservate si pone:

Definizione 17. *Dato uno spazio di probabilità* (S, P)*, ed una variabile aleatoria discreta* X *si dice valore atteso, o speranza matematica, di* X *il valore*

$$E(X) = \sum_{s \in S} X(s) P(s)$$

se

$$E(|X|) = \sum_{s \in S} |X(s)| P(s) < \infty;$$

in altre parole si richiede la convergenza assoluta della serie che definisce il valore atteso.

Esempio 50. *Per il caso (b) dell'esempio precedente il valore atteso non esiste e si dovranno sviluppare altri metodi.*

Esempio 51. *Se* $X \sim$ *geometrica(p) allora* $X \geq 0$ *e quindi basta che sia finita* $E(X)$ *stessa: derivando per serie, come lecito all'interno del raggio di convergenza di una serie di potenze, si ha:*

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} kp(1-p)^{k-1} \qquad (5.18) \\ &= -\sum_{k=1}^{\infty} p \frac{d}{dp} (1-p)^k \\ &= -\frac{d}{dp} \sum_{k=1}^{\infty} p(1-p)^k \\ &= -p \frac{d}{dp} \frac{1}{p} = \frac{1}{p}. \end{aligned}$$

Intuitivamente, se la probabilità di successo è $p = \frac{1}{m}$, allora il primo successo arriverà in media alla m-sima prova.

Esempio 52. *Se* $X \sim$ *Poisson*(λ) *allora* $X \geq 0$ *e*

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \qquad (5.19) \\ &= \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda. \end{aligned}$$

Avendo quindi assunto la convergenza assoluta della serie che definisce il valore atteso, valgono tutti i risultati relativi alle proprietà del valore atteso. In particolare valgono i Lemmi 6 e 7. Il lettore è invitato a verificare che le dimostrazioni dei Lemmi suddetti possono essere adattate anche al caso presente.

Come esempio mostriamo come si adatta la dimostrazione della parte (ii) del Lemma 7 nel caso $a_1 = a_2 = 1, X_1 = X, X_2 = Y$. Assumendo che $E(X)$ ed $E(Y)$ esistano, prima si dimostra che $E(X+Y)$ esiste e poi se ne calcola il valore. Per la diseguaglianza triangolare, poichè $X \leq Y$ implica $E(X) \leq E(Y)$ e poiché si può cambiare a piacere l'ordine di sommazione di una serie a termini positivi convergente, si ha:

$$\begin{aligned} E(|X+Y|) &\leq E(|X|+|Y|) \\ &= \sum_{s \in S} (|X(s)| + |Y(s)|) P(s) \\ &= \sum_{s \in S} (|X(s)| P(s) + \sum_{s \in S} (|Y(s)| P(s) \\ &= E(|X|) + E(|Y|) < \infty \qquad (5.20) \end{aligned}$$

dall'ipotesi; la finitezza del risultato giustifica a posteriori la riorganizzazione delle somme. Ora

$$\begin{aligned} E(X+Y) &= \sum_{s \in S} (X(s) + Y(s)) P(s) \\ &= \sum_{s \in S} (X(s) P(s) + \sum_{s \in S} (Y(s) P(s) \\ &= E(X) + E(Y) < \infty \qquad (5.21) \end{aligned}$$

in cui abbiamo riorganizzato nuovamente l'ordine di sommazione in quanto le serie coinvolte sono assolutamente convergenti.

Esercizio 14. *Verificare che le altre dimostrazioni citate si estendono al caso delle variabili discrete.*

Aggiungiamo un'altra semplice conseguenza della diseguaglianza triangolare:

Esercizio 15. *Mostrare che* $|E(X)| \leq E(|X|)$.

Considerando le difficoltà relative ai prodotti infiniti la definizione di indipendenza di una famiglia di variabili aleatorie discrete può essere data riferendosi a sottofamiglie finite:

Definizione 18. *Le variabili di una famiglia al più numerabile di variabili aleatorie* $X_1, \dots X_n, \dots$ *si dicono indipendenti se sono indipendenti le variabili aleatorie in ogni sottofamiglia finita.*

Esercizio 16. *Verificare che valgono il Lemma 8 ed il teorema 7 nel caso delle variabili discrete.*

Anche per le variabili aleatorie discrete il valore atteso del prodotto di variabili indipendenti è uguale al prodotto dei valori attesi, questa volta però occorre introdurre una condizione sufficiente affinché il valore atteso del prodotto esista. Per questo bisogna conviene premettere la discussione sul valore atteso di funzioni delle variabili aleatorie.

Infatti, se ϕ è una funzione definita su S_X allora $\phi(X)$ è una variabile aleatoria discreta, ma non è detto che anche se X ha valore atteso ci sia il valore atteso di $\phi(X)$.

Esercizio 17. *Mostrare che esistono una variabile aleatoria* X *ed una funzione* ϕ *definita su* S_X *tali che* $E(X)$ *esiste ma* $E(\phi(X))$ *non esiste.*

Tuttavia, se il valore atteso di $\phi(X)$ esiste, allora si può calcolare con il cambiamento di variabili:

Esercizio 18. *Verificare che se* X *è una variabile aleatoria discreta e* ϕ *è una funzione definita su* S_X *e se* $E(\phi(X))$ *ammette valore atteso, allora vale il Teorema 6.*

Ora torniamo alla questione del valore atteso di variabili aleatorie indipendenti. Analogamente a prima, dalla sola esistenza del valore atteso di X ed Y non si può dedurre l'esistenza del valore atteso di XY o viceversa.

Esercizio 19. *Mostrare che esistono variabili aleatorie* X *ed* Y *tali che* $E(X)$ *ed* $E(Y)$ *esistono ma non esiste* $E(XY)$*. Viceversa, mostrare che esistono variabili aleatorie* X *ed* Y *tali che* $E(XY)$ *esiste ma* $E(X)$ *ed* $E(Y)$ *non esistono.*

Occorre quindi condizioni per garantire l'esistenza di altri valori attesi. Un primo risultato dice che i momenti successivi implicano l'esistenza dei momenti precedenti.

Lemma 13. *Se* X *è una variabile aleatoria discreta tale che* $E(X^k)$ *esiste, con questo intendendo che* $X \geq 0$ *oppure* $k \in \mathbb{N}$*, per qualche* k*, allora esiste* $E(X^h)$ *per ogni* $h \leq k$*, sempre intendendo* $X \geq 0$ *oppure* h *intero.*

DIMOSTRAZIONE.

$$\begin{aligned}
E(|X^h|) &= \sum_{s\in S} |X(s)^h| P(s) \\
&= \sum_{s\in S:|X(s)^h|\leq 1} |X(s)^h| P(s) + \sum_{s\in S:|X(s)^h|>1} |X(s)^h| P(s) \\
&\leq P(|X^h| \leq 1) + \sum_{s\in S:|X(s)^h|>1} |X(s)^k| P(s) \\
&\leq P(|X^h| \leq 1) + \sum_{s\in S} |X(s)^k| P(s) < \infty
\end{aligned}$$

□

LEMMA 14. *Se X ed Y sono variabili aleatorie discrete tali che $E(X^2)$ ed $E(Y^2)$ esistono, allora esistono $E(X)$, $E(Y)$ ed $E(XY)$.*

DIMOSTRAZIONE. L'esistenza di $E(X)$ ed $E(Y)$ segue dal lemma precedente. Da $0 \leq (a-b)^2 = a^2 + b^2 - 2ab$ si ha che per qualsiasi coppia di numeri reali a e b vale che $ab \leq (a^2+b^2)/2$, per cui

$$\begin{aligned}
E(|XY|) &= \sum_{s\in S} |X(s)||Y(s)| P(s) \\
&= \frac{1}{2}\sum_{s\in S}(|X(s)| + |Y(s)|)P(s) \\
&\leq \frac{1}{2}(\sum_{s\in S}|X(s)|P(s) + \sum_{s\in S}|Y(s)|P(s)) < \infty
\end{aligned}$$

□

Ora abbiamo condizioni sufficienti per generalizzare i risultati sui valori attesi di variabili indipendenti.

ESERCIZIO 20. *Verificare che se X_i sono variabili aleatorie discrete tali che $E(X_i^2)$ esiste per ogni i allora vale il Teorema 8.*

Con il secondo momento è quindi possibile definire la varianza e la deviazione standard di X e vale anche in questo caso l'additività delle varianze per variabili indipendenti.

ESERCIZIO 21. *Verificare che se X è una variabile aleatoria discreta tale che $E(X^2)$ esiste allora la varianza e la deviazione standard di X esistono e vale il Teorema 9.*