1.3.1

Proposizione 1.3.1 *Sia $(\Omega, \mathscr{F}, P)$ uno spazio di probabilità. Allora:*

1. *se $E \in \mathscr{F}$ allora $P(E^c) = 1 - P(E)$ (*probabilità del complementare*);*
2. *se $E \in \mathscr{F}$ allora $P(E) \leq 1$;*

3. *se $E, F \in \mathscr{F}$ e $F \subset E$ allora $P(E \setminus F) = P(E) - P(F)$;*
4. *se $E, F \in \mathscr{F}$ e $F \subset E$ allora $P(F) \leq P(E)$ (*monotonia*);*
5. *se $E, F \in \mathscr{F}$ allora $P(E \cup F) = P(E) + P(F) - P(E \cap F)$ (*probabilità dell'unione*).*

Dimostrazione

1. Notiamo che $\Omega = E \cup E^c$ e $E \cap E^c = \emptyset$; quindi per l'assioma 2. della Definizione 1.2.16 e il punto 2. della Proposizione 1.2.17 vale $1 = P(\Omega) = P(E) + P(E^c)$ che implica $P(E^c) = 1 - P(E)$.

2. Per il punto precedente $P(E) = 1 - P(E^c)$, ma $P(E^c) \geq 0$ per l'assioma 1. della Definizione 1.2.16; segue che necessariamente $P(E) \leq 1$.

3. Se $F \subset E$ allora $E = (E \setminus F) \cup F$ e l'unione è disgiunta; applicando il punto 2. della Proposizione 1.2.17: $P(E) = P(E \setminus F) + P(F)$ e quindi $P(E \setminus F) = P(E) - P(F)$.

4. Per il punto precedente $P(E) - P(F) = P(E \setminus F)$ che è non negativo per l'assioma 1. della Definizione 1.2.16.

5. Possiamo scrivere $E \cup F = (E \cap F^c) \cup (E \cap F) \cup (E^c \cap F)$ e l'unione è disgiunta; sempre il punto 2. della Proposizione 1.2.17 implica $P(E \cup F) = P(E \cap F^c) + P(E \cap F) + P(E^c \cap F)$; quindi $P(E \cup F) + P(E \cap F) = P(E \cap F^c) + P(E \cap F) + P(E \cap F) + P(E^c \cap F)$. Ma $P(E \cap F^c) + P(E \cap F) = P(E)$ e $P(E \cap F) + P(E^c \cap F) = P(F)$ (verificare!) e quindi $P(E \cup F) + P(E \cap F) = P(E) + P(F)$. ∎

Applicando due volte la proprietà 5. della Proposizione 1.3.1, possiamo calcolare la probabilità dell'unione di tre eventi $E, F, G \in \mathscr{F}$:

$$\begin{aligned} P(E \cup F \cup G) &= P((E \cup F) \cup G) \\ &= [P(E) + P(F) + P(G)] - [P(E \cap F) + P(E \cap G) + P(F \cap G)] + P(E \cap F \cap G) \end{aligned}$$

1.5.9

Proposizione 1.5.9 (Formula delle probabilità totali) *Sia* $(\Omega, \mathscr{F}, P)$ *uno spazio di probabilità e* $F_1, F_2, \ldots, F_n \in \mathscr{F}$ *una* partizione finita di Ω, $\bigcup_{k=1}^n F_k = \Omega$ *e* $F_h \cap F_k = \emptyset$ *se* $h \neq k$, *tale che* $P(F_k) > 0$ *per* $k = 1, 2, \ldots, n$. *Allora per ogni evento* $E \in \mathscr{F}$ *si ha*

$$P(E) = \sum_{k=1}^{n} P(E|F_k)P(F_k) \tag{1.5.1}$$

Dimostrazione Sia $E \in \mathscr{F}$, poiché $\Omega = \bigcup_{k=1}^n F_k$ ed $E \subset \Omega$, segue che $E = E \cap \Omega = \bigcup_{k=1}^n (E \cap F_k)$; inoltre, poiché $F_h \cap F_k = \emptyset$ se $h \neq k$, allora $\bigcup_{k=1}^n (E \cap F_k)$ è un'unione disgiunta e dall'additività otteniamo

$$P(E) = \sum_{k=1}^{n} P(E \cap F_k) = \sum_{k=1}^{n} P(E|F_k)P(F_k)$$

(l'ultima uguaglianza segue direttamente dalla definizione di probabilità condizionata). ∎

1.5.13

Proposizione 1.5.13 (Formula di Bayes) *Sia* $(\Omega, \mathscr{F}, P)$ *uno spazio di probabilità e* $F_1, F_2, \ldots, F_n \in \mathscr{F}$ *una partizione finita di* Ω *tale che* $P(F_k) > 0$ *per* $k = 1, 2, \ldots, n$. *Se* $E \in \mathscr{F}$ *è tale che* $P(E) > 0$ *allora si ha*

$$P(F_h|E) = \frac{P(E|F_h)P(F_h)}{\sum_{k=1}^{n} P(E|F_k)P(F_k)} \qquad h = 1, 2, \ldots, n \tag{1.5.2}$$

Dimostrazione Dalla definizione di probabilità condizionata si ha

$$P(F_h|E) = \frac{P(F_h \cap E)}{P(E)} = \frac{P(E|F_h)P(F_h)}{P(E)}$$

cosi che la (1.5.2) si ottiene applicando la formula delle probabilità totali (1.5.1) al denominatore di questa uguaglianza. ∎

1.5.36

Proposizione 1.5.36 *La probabilità di osservare $k \leq n$ successi in una sequenza di $n \geq 1$ prove di Bernoulli se la probabilità di successo della singola prova è $p \in (0,1)$ è data da*

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Dimostrazione Sia $(\Omega, \mathscr{F}, P)$ lo spazio di probabilità di Bernoulli, e $B_k \in \mathscr{F}$ l'evento "si osservano k successi", cioè

$$B_k = \left\{ (a_1, a_2, \ldots, a_n) \in \Omega \colon \sum_{h=1}^{n} a_h = k \right\}$$

allora

$$P(B_k) = \sum_{\omega \in B_k} P(\{\omega\}) = \sum_{\omega \in B_k} p^k (1-p)^{n-k} = |B_k| p^k (1-p)^{n-k}$$

ma $|B_k| = \binom{n}{k}$, infatti per elencare tutte le stringhe lunghe n in cui k cifre sono uguali ad 1 ed $n-k$ sono uguali a 0, basta fissare i k posti degli 1 e questo può essere fatto in $\binom{n}{k}$ modi. ∎

Risultati sezione 2.3.3

$$P(X=k)=\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

Per capire cosa succede a $P(X=k)$ se n è grande osserviamo che

$$\begin{aligned}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}&=\frac{n!}{(n-k)!k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\&=\frac{n!}{(n-k)!n^k}\left(1-\frac{\lambda}{n}\right)^{-k}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n}\end{aligned}$$

ma

$$\lim_{n\to+\infty}\frac{n!}{(n-k)!n^k}=1$$

come rapporto di polinomi di grado k,

$$\lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1$$

e

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=\mathrm{e}^{-\lambda}$$

come ben noto dal corso di analisi. Segue che

$$P(X=k)\simeq\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\qquad k=0,1\dots,n,\qquad \lambda=np \tag{2.3.1}$$

Tenendo conto di quanto detto sopra, per $\lambda>0$ introduciamo la densità

$$p(k):=\begin{cases}\dfrac{\mathrm{e}^{-\lambda}\lambda^k}{k!} & \text{se } k\in\{0,1,2,\dots\}\\ 0 & \text{se } k\notin\{0,1,2,\dots\}\end{cases}$$

che prende il nome di *densità di Poisson di parametro* λ. Una variabile aleatoria con questa densità è detta *variabile di Poisson di parametro* λ e si scrive $X\sim\mathcal{P}(\lambda)$.

3.2.5

Proposizione 3.2.5 *Sia X una variabile aleatoria, allora*

1. $\operatorname{Var}(X) = 0$ *se e solo se* $P(X = c) = 1$ *per qualche costante* c. *In questo caso* $c = \mathrm{E}(X)$.

2. *Se X ammette varianza ed* $\alpha \in \mathbb{R}$ *allora* $\operatorname{Var}(\alpha X) = \alpha^2 \operatorname{Var}(X)$.

3. *Se X ammette varianza e* $\beta \in \mathbb{R}$ *allora* $\operatorname{Var}(X + \beta) = \operatorname{Var}(X)$.

4. *Se X ammette varianza allora X^2 ammette media e* $\operatorname{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$.

Dimostrazione Utilizzeremo le proprietà della media contenute nella Proposizione 3.1.14.

1. Se $P(X = c) = 1$ allora $\mathrm{E}(X) = c$ e $\operatorname{Var}(X) = \mathrm{E}((c - c)^2) = \mathrm{E}(0) = 0$. Viceversa, se $\operatorname{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2) = 0$, poiché $P(X - \mathrm{E}(X))^2 \geq 0) = 1$ allora $P((X - \mathrm{E}(X))^2 = 0) = 1$ che è possibile solo se $P(X = \mathrm{E}(X)) = 1$.

2. Poiché $\mathrm{E}(\alpha X) = \alpha \mathrm{E}(X)$ allora
$$\begin{aligned}\operatorname{Var}(\alpha X) &= \mathrm{E}((\alpha X - \mathrm{E}(\alpha X))^2) = \mathrm{E}((\alpha X - \alpha \mathrm{E}(X))^2) \\ &= \mathrm{E}(\alpha^2 (X - \mathrm{E}(X))^2) = \alpha^2 \mathrm{E}((X - \mathrm{E}(X))^2) = \alpha^2 \operatorname{Var}(X).\end{aligned}$$

3. Osserviamo che per la linearità del valore atteso $\mathrm{E}(X + \beta) = \mathrm{E}(X) + \beta$, quindi
$$\begin{aligned}\operatorname{Var}(X + \beta) &= \mathrm{E}((X + \beta - \mathrm{E}(X + \beta))^2) \\ &= \mathrm{E}((X + \beta - \mathrm{E}(X) - \beta)^2) \\ &= \mathrm{E}((X - \mathrm{E}(X))^2) \\ &= \operatorname{Var}(X).\end{aligned}$$

4. Se X ammette varianza allora
$$\begin{aligned}\mathrm{E}(X^2) &= \mathrm{E}((X - \mathrm{E}(X) + \mathrm{E}(X))^2) \\ &\leq \mathrm{E}(2(X - \mathrm{E}(X))^2 + 2\,\mathrm{E}(X)^2) \\ &= 2\,\mathrm{E}((X - \mathrm{E}(X))^2) + 2\,\mathrm{E}(X)^2 \\ &= 2 \operatorname{Var}(X) + 2\,\mathrm{E}(X)^2 < +\infty\end{aligned}$$

   Quindi X^2 ammette media. Inoltre:
$$\begin{aligned}\operatorname{Var}(X) &= \mathrm{E}((X - \mathrm{E}(X))^2) \\ &= \mathrm{E}(X^2 - 2X\,\mathrm{E}(X) + \mathrm{E}(X)^2) \\ &= \mathrm{E}(X^2) - 2\,\mathrm{E}(X\,\mathrm{E}(X)) + \mathrm{E}(\mathrm{E}(X)^2) \\ &= \mathrm{E}(X^2) - 2\,\mathrm{E}(X)^2 + \mathrm{E}(X)^2 \\ &= \mathrm{E}(X^2) - \mathrm{E}(X)^2 \qquad \blacksquare\end{aligned}$$

Commentiamo brevemente la proposizione appena dimostrata. Il punto 1. afferma che le uniche variabili aleatorie con varianza nulla sono le costanti. Questo è in pieno accordo con il concetto intuitivo di varianza come misura di quanto una variabile aleatoria si discosta dalla propria media. Il punto 2. ci dice che la varianza è quadratica (mentre la media è lineare). Il punto 3. mostra come la varianza sia invariante per traslazioni. Infatti sommando ad una variabile aleatoria un numero, cioè traslandola, anche la media viene traslata dello stesso numero e lo scostamento della variabile dalla sua media non cambia. Il punto 4. mostra una formula molto utile nelle applicazioni e negli esercizi per calcolare la varianza.

A titolo d'esempio calcoliamo la varianza di alcune delle variabili aleatorie precedentemente introdotte.

3.3.1

3.3 Disuguaglianza di Chebychev

La successiva importante disuguaglianza, nota come *disuguaglianza di Chebychev*, precisa in che senso una variabile X con varianza "piccola" è concentrata intorno alla sua media.

Proposizione 3.3.1 (Disuguaglianza di Chebychev) *Sia X una variabile aleatoria che ammette media e varianza. Allora per ogni $\epsilon > 0$:*

$$P(|X - \mathrm{E}(X)| > \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}$$

Dimostrazione Osserviamo che o $|X - \mathrm{E}(X)| \leq \epsilon$ oppure $|X - \mathrm{E}(X)| > \epsilon$; quindi $\mathbf{1}_{(-\infty,\epsilon]}(|X - \mathrm{E}(X)|) + \mathbf{1}_{(\epsilon,+\infty)}(|X - \mathrm{E}(X)|) \equiv 1$, da cui

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}((X - \mathrm{E}(X))^2) \\
&= \mathrm{E}((X - \mathrm{E}(X))^2 \mathbf{1}_{(-\infty,\epsilon]}(|X - \mathrm{E}(X)|)) + \mathrm{E}((X - \mathrm{E}(X))^2 \mathbf{1}_{(\epsilon,+\infty)}(|X - \mathrm{E}(X)|) \\
&\geq \mathrm{E}((X - \mathrm{E}(X))^2 \mathbf{1}_{(\epsilon,+\infty)}(|X - \mathrm{E}(X)|) \\
&\geq \mathrm{E}(\epsilon^2 \mathbf{1}_{(\epsilon,+\infty)}(|X - \mathrm{E}(X)|)) \\
&= \epsilon^2 P(|X - \mathrm{E}(X)| > \epsilon) \qquad \blacksquare
\end{aligned}$$

3.3.2

Esercizio 3.3.2 Dimostrare che se X è una variabile aleatoria positiva tale che la k-esima potenza X^k ammette media per un intero positivo k, allora vale che

$$P(X > \epsilon) \leq \frac{\mathrm{E}(X^k)}{\epsilon^k} \qquad \forall \epsilon > 0.$$

Questa disuguaglianza è nota con il nome di *Disuguaglianza di Markov*.

4.4.3

Proposizione 4.4.3 *Se $f_{\boldsymbol{X}}$ è la densità di un vettore aleatorio n–dimensionale assolutamente continuo $\boldsymbol{X} = (X_1, \ldots, X_n)$ allora X_i è una variabile aleatoria assolutamente continua e la sua densità è*

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f_{\boldsymbol{X}}(s_1, \ldots, s_{i-1}, x_i, s_{i+1}, \ldots, s_n)\, ds_1 \cdots ds_{i-1}\, ds_{i+1} \cdots ds_n$$

Dimostrazione Per semplicità di notazioni, consideriamo il caso $i = 1$. Bisogna dimostrare che

$$F_{X_1}(x) = \int_{-\infty}^{x} \left\{ \int_{\mathbb{R}^{n-1}} f_{\boldsymbol{X}}(s_1, \ldots, s_n)\, ds_2 \cdots ds_n \right\} ds_1$$

che è vero in quanto, se $B := (-\infty, x] \times \mathbb{R}^{n-1}$, allora per il punto 3. della Proposizione 4.4.2 abbiamo:

$$\int_{-\infty}^{x} \left\{ \int_{\mathbb{R}^{n-1}} f_{\boldsymbol{X}}(s_1, \cdots, s_n)\, ds_2 \ldots ds_n \right\} ds_1 = P(\boldsymbol{X} \in B) =$$
$$= P(X_1 \le x, X_2 \in \mathbb{R}, \ldots, X_n \in \mathbb{R}) = P(X_1 \le x) = F_{X_1}(x) \quad \blacksquare$$

Le densità delle componenti di un vettore assolutamente continuo sono dette *densità marginali.*

Somme di variabili aleatorie pag.90

Somme di variabili aleatorie: caso di un vettore assolutamente continuo. Sia (X_1, X_2) un vettore aleatorio assolutamente continuo con densità f_{X_1,X_2}. L'equazione (4.5.2) applicata alla funzione $g(x_1, x_2) = x_1+x_2$ fornisce per la funzione di ripartizione di X_1+X_2:

$$\begin{aligned}
F_{X_1+X_2}(y) &= \int_{\{(x_1,x_2):\, x_1+x_2\le y\}} f_{X_1,X_2}(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{y-x_1} f_{X_1,X_2}(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_{-\infty}^{y}\left(\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2 - x_1)\, dx_1\right) dx_2
\end{aligned}$$

Quindi, $X_1 + X_2$ è una variabile aleatoria assolutamente continua e ha densità

$$f_{X_1+X_2}(y) = \int_{-\infty}^{+\infty} f_{X_1,X_2}(x_1, y - x_1)\; dx_1$$

Inoltre, se X_1, X_2 sono indipendenti allora

$$f_{X_1+X_2}(y) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) f_{X_2}(y - x_1)\; dx_1 \tag{4.5.3}$$

Calcoliamo ora le densità delle somme di alcune variabili aleatorie indipendenti assolutamente continue.

Esempio 4.5.5 (Somme di variabili aleatorie gaussiane indipendenti) Cominciamo sommando due variabili aleatorie Z_1, Z_2 gaussiane indipendenti e a media nulla, cioè $Z_1 \sim \mathcal{N}(0, \sigma_1^2)$ e $Z_2 \sim \mathcal{N}(0, \sigma_2^2)$. Segue dalla (4.5.3) che

$$\begin{aligned}
f_{Z_1+Z_2}(y) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2} \mathrm{e}^{-\frac{x^2}{2\sigma_1^2}-\frac{(y-x)^2}{2\sigma_2^2}}\; dx \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \mathrm{e}^{-\frac{y^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\tau}} \mathrm{e}^{-\frac{(x-\nu)^2}{2\tau}}\; dx
\end{aligned}$$

dove

$$\nu := \frac{y\sigma_1^2}{\sigma_1^2+\sigma_2^2} \quad \text{e} \quad \tau := \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

D'altro canto

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\tau}} \mathrm{e}^{-\frac{(x-\nu)^2}{2\tau}}\, dx = 1$$

Quindi

$$f_{Z_1+Z_2}(y) = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \mathrm{e}^{-\frac{y^2}{2(\sigma_1^2+\sigma_2^2)}}$$

cioè $Z_1 + Z_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

Siano ora X_1, X_2 due variabili aleatorie indipendenti con $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Allora $X_1 + X_2$ ha la stessa densità di $(Z_1 + Z_2) + (\mu_1 + \mu_2)$, che è trasformazione lineare della variabile aleatoria gaussiana $Z_1 + Z_2$, come abbiamo appena dimostrato. Quindi: $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Iterando il procedimento ora visto otteniamo che se $X_1, \ldots, X_n$ sono variabili aleatorie indipendenti e gaussiane con $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $\forall\ i = 1, \ldots, n$, allora $\sum_{i=1}^n X_i \sim \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$. In breve: la somma di variabili aleatorie gaussiane indipendenti è gaussiana di parametri la somma dei parametri.

Esempio 4.5.6 Siano X_1, X_2 due variabili aleatorie indipendenti entrambe con densità esponenziale di parametro $\lambda > 0$. Calcoliamo la densità di $X_1 + X_2$.
Applicando (4.5.3) abbiamo:

$$f_{X_1+X_2}(y) = \begin{cases} \int_0^y \lambda \mathrm{e}^{-\lambda u} \lambda \mathrm{e}^{-\lambda(y-u)}\, du = \lambda^2 \mathrm{e}^{-\lambda y} y & \text{se } y > 0 \\ 0 & \text{se } y \leq 0 \end{cases}$$

Procedendo per induzione su n si può dimostrare che se $X_1, \ldots, X_n$ sono variabili aleatorie indipendenti tali che $X_i \sim \mathcal{E}(\lambda)$ $\forall\ i = 1, \ldots, n$, allora la densità di $X_1 + \cdots + X_n$ è

$$f_{X_1+\cdots+X_n}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} \mathrm{e}^{-\lambda x} \mathbf{1}_{(0,+\infty)}(x) \tag{4.5.4}$$

Definizione 4.5.7 *La densità* (4.5.4) *è detta densità* Gamma di parametri n e λ *e scriveremo* $\Gamma(n, \lambda)$.

4.7.4

Corollario 4.7.4 *Siano X_1 e X_2 variabili aleatorie indipendenti e che ammettono media. Allora anche X_1X_2 ammette media e*

$$\mathrm{E}(X_1X_2) = \mathrm{E}(X_1)\,\mathrm{E}(X_2)$$

Dimostrazione Supponiamo che X_1 e X_2 siano assolutamente continue con densità rispettivamente f_{X_1} e f_{X_2}. Allora:

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |xy| f_{X_1}(x) f_{X_2}(y)\, dx\, dy = \int_{\mathbb{R}} |x| f_{X_1}(x)\, dx \cdot \int_{\mathbb{R}} |y| f_{X_2}(y)\, dy < +\infty$$

e $\mathrm{E}(X_1X_2)$ esiste per la Proposizione 4.7.1 applicata alla funzione $g(x,y) = xy$. Inoltre, dalla Proposizione 4.7.1 discende che:

$$\mathrm{E}(X_1X_2) = \int_{\mathbb{R}}\int_{\mathbb{R}} xy f_{X_1}(x) f_{X_2}(y)\, dx\, dy = \int_{\mathbb{R}} x f_{X_1}(x)\, dx \cdot \int_{\mathbb{R}} y f_{X_2}(y)\, dy = \mathrm{E}(X_1)\,\mathrm{E}(X_2) \quad \blacksquare$$

4.7.6

Corollario 4.7.6 *Se X_1 e X_2 hanno varianza (finita), rispettivamente $\mathrm{Var}(X_1)$ e $\mathrm{Var}(X_2)$, allora anche $X_1 + X_2$ ha varianza finita e*

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{E}[(X_1 - \mathrm{E}(X_1))(X_2 - \mathrm{E}(X_2))] \tag{4.7.4}$$

Inoltre, se X_1, X_2 sono indipendenti allora

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) \tag{4.7.5}$$

Dimostrazione Poiché $((X_1 + X_2) - \mathrm{E}(X_1 + X_2))^2 = [(X_1 - \mathrm{E}(X_1)) + (X_2 - \mathrm{E}(X_2))]^2 \leq 2[(X_1 - \mathrm{E}(X_1))^2 + (X_2 - \mathrm{E}(X_2))^2]$, allora $\mathrm{Var}(X_1 + X_2) = \mathrm{E}[((X_1 + X_2) - \mathrm{E}(X_1 + X_2))^2] = \mathrm{E}[((X_1 - \mathrm{E}(X_1)) + (X_2 - \mathrm{E}(X_2)))^2] \leq 2[\mathrm{E}(X_1 - \mathrm{E}(X_1))^2 + \mathrm{E}(X_2 - \mathrm{E}(X_2))^2] = 2(\mathrm{Var}(X_1) + \mathrm{Var}(X_2))$. Quindi se X_1 e X_2 ammettono varianza, anche $X_1 + X_2$ la ammette. Inoltre

$$\begin{aligned}
\mathrm{Var}(X_1 + X_2) &= \mathrm{E}[((X_1 - \mathrm{E}(X_1)) + (X_2 - \mathrm{E}(X_2)))^2] \\
&= \mathrm{E}[(X_1 - \mathrm{E}(X_1))^2 + (X_2 - \mathrm{E}(X_2))^2 + 2(X_1 - \mathrm{E}(X_1))(X_2 - \mathrm{E}(X_2))] \\
&= \mathrm{E}[(X_1 - \mathrm{E}(X_1))^2] + \mathrm{E}[(X_2 - \mathrm{E}(X_2))^2] + 2\,\mathrm{E}[(X_1 - \mathrm{E}(X_1))(X_2 - \mathrm{E}(X_2))]
\end{aligned}$$

[dove l'ultima eguaglianza deriva dal Corollario 4.7.2 applicato alla somma delle variabili $(X_1 - \mathrm{E}(X_1))^2$, $(X_2 - \mathrm{E}(X_2))^2$ e $(X_1 - \mathrm{E}(X_1))(X_2 - \mathrm{E}(X_2))$]

$$= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{E}[(X_1 - \mathrm{E}(X_1))(X_2 - \mathrm{E}(X_2))]$$

Per completare la dimostrazione, basta notare che se X_1, X_2 sono indipendenti, allora, per la Proposizione 4.6.2, anche $X_1 - \mathrm{E}(X_1), X_2 - \mathrm{E}(X_2)$ sono indipendenti e $\mathrm{E}[(X_1 - \mathrm{E}(X_1))(X_2 - \mathrm{E}(X_2))] = 0$ in virtù del Corollario 4.7.4. $\blacksquare$

4.8.2

Proposizione 4.8.2 *Siano X_1, X_2, X_3 variabili aleatorie con varianza finita e $a, b \in \mathbb{R}$. Allora*

1. $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1)$;
2. $\mathrm{Cov}(aX_1, X_2) = a\,\mathrm{Cov}(X_1, X_2)$;
3. $\mathrm{Cov}(X_1 + X_2, X_3) = \mathrm{Cov}(X_1, X_3) + \mathrm{Cov}(X_2, X_3)$;
4. $\mathrm{Cov}(X_1, X_2) = \mathrm{E}(X_1X_2) - \mathrm{E}(X_1)\,\mathrm{E}(X_2)$;
5. *se X_1, X_2 sono indipendenti allora* $\mathrm{Cov}(X_1, X_2) = 0$;
6. $|\rho_{X_1,X_2}| \leq 1$ *e* $|\rho_{X_1,X_2}| = 1$ *se e solo se esistono $a, b \in \mathbb{R}$ tali che $P(X_2 = aX_1 + b) = 1$. Inoltre in tal caso:*
$$a = \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)} \qquad e \qquad b = \mathrm{E}(X_2) - \frac{\mathrm{E}(X_1)\,\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}$$

Dimostrazione Le proprietà 1.–5. seguono immediatamente dalle proprietà della media e la dimostrazione viene lasciata per esercizio al lettore.

La dimostrazione della proprietà 6. è mutuata da [7], pag. 329 e si basa sulle proprietà della varianza. Siano σ_1^2, σ_2^2 le varianze di X_1, X_2, rispettivamente. Allora

$$\begin{aligned}
0 \leq \mathrm{Var}\left(\frac{X_1}{\sigma_1} + \frac{X_2}{\sigma_2}\right) &= \frac{\mathrm{Var}(X_1)}{\sigma_1^2} + \frac{\mathrm{Var}(X_2)}{\sigma_2^2} + 2\,\mathrm{Cov}\left(\frac{X_1}{\sigma_1}, \frac{X_2}{\sigma_2}\right) \\
&= \frac{\sigma_1^2}{\sigma_1^2} + \frac{\sigma_2^2}{\sigma_2^2} + 2\frac{\mathrm{Cov}\,(X_1, X_2)}{\sigma_1\sigma_2} \quad \text{[per il punto 2.]} \\
&= 2(1 + \rho_{X_1,X_2})
\end{aligned}$$

da cui otteniamo

$$\rho_{X_1,X_2} \geq -1$$

Inoltre,

$$0 \leq \mathrm{Var}\left(\frac{X_1}{\sigma_1} - \frac{X_2}{\sigma_2}\right) = \frac{\mathrm{Var}(X_1)}{\sigma_1^2} + \frac{\mathrm{Var}(X_2)}{\sigma_2^2} - 2\frac{\mathrm{Cov}\,(X_1, X_2)}{\sigma_1\sigma_2} = 2(1 - \rho_{X_1,X_2})$$

e quindi

$$\rho_{X_1,X_2} \leq 1$$

Per dimostrare la seconda parte della proprietà 6., osserviamo che $\rho_{X_1,X_2} = 1$ se e solo se $\mathrm{Var}\,(X_1/\sigma_1 - X_2/\sigma_2) = 0$. Segue quindi dalle proprietà della varianza che

$$\rho_{X_1,X_2} = 1 \text{ se e solo se } P\left(\frac{X_1}{\sigma_1} - \frac{X_2}{\sigma_2} = \frac{\mathrm{E}(X_1)}{\sigma_1} - \frac{\mathrm{E}(X_2)}{\sigma_2}\right) = 1$$

Inoltre, $\rho_{X_1,X_2} = 1$ se e solo se $\mathrm{Cov}(X_1, X_2) = \sigma_1\sigma_2$ e quindi $\rho_{X_1,X_2} = 1$ se e solo se

$$X_2 = \mathrm{E}(X_2) + \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_1^2}(X_1 - \mathrm{E}(X_1))$$

Invece, per $\rho_{X_1,X_2} = -1$ valgono le seguenti equivalenze che completano la dimostrazione: $\rho_{X_1,X_2} = -1$ se e solo se $\mathrm{Cov}(X_1, X_2) = -\sigma_1\sigma_2$ se e solo se $\mathrm{Var}\,(X_1/\sigma_1 + X_2/\sigma_2) = 0$ se e solo se $P\,(X_1/\sigma_1 + X_2/\sigma_2 = \mathrm{E}(X_1)/\sigma_1 + \mathrm{E}(X_2)/\sigma_2) = 1$ se e solo se

$$X_2 = \mathrm{E}(X_2) + \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_1^2}(X_1 - \mathrm{E}(X_1)) \quad \blacksquare$$

4.11.1

Proposizione 4.11.1 (Legge debole dei grandi numeri) *Sia $X_1, X_2, \dots$ una successione di variabili aleatorie indipendenti ed identicamente distribuite (i.i.d.) con media μ e varianza σ^2 finite. Sia $S_n = X_1 + \cdots + X_n$. Allora, per ogni $\epsilon > 0$*

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) = 0$$

Dimostrazione Poiché le X_i sono i.i.d. allora

$$\mathrm{Var}(S_n) = n\,\mathrm{Var}(X_1) = n\sigma^2$$

da cui

$$\mathrm{Var}(\frac{S_n}{n}) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

e

$$\mathrm{E}(\frac{S_n}{n}) = \mu.$$

Segue dalla diseguaglianza di Chebychev che per ogni $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \quad (n \to +\infty) \quad \blacksquare$$

Date n variabili aleatorie $X_1, \dots, X_n$ si chiama *media campionaria* di $X_1, \dots, X_n$ la quantità $(X_1 + \cdots + X_n)/n$ e la si indica con $\overline{X}_n$. Equivalentemente, la Legge debole dei grandi numeri afferma che $P\left(\left|\overline{X}_n - \mu\right| \leq \epsilon\right) \to 1$ per $n \to +\infty$; quindi, essa mette in evidenza che, pur partendo da un esperimento aleatorio costituito da prove ripetute del quale poco si può predire ad ogni prova (le prove sono indipendenti), facendo le medie di tali prove si ottiene un esperimento il cui risultato può essere predetto con un elevato grado di certezza. In realtà vale il seguente risultato "più forte" la cui dimostrazione è più laboriosa.

TCL

2. Dal Teorema Centrale del Limite (enunciato: Teorema 4.11.6 nelle dispense) abbiamo che, per v.a. i.i.d. con varianza finita, la media campionaria è asintoticamente normale di media asintotica $E_p[X_i] = p$ e varianza asintotica $\mathrm{Var}_p[X_i]/n = p(1-p)/n$. ▮

Def. errore quadratico MSE e sua decomposizione con dimostrazione (2 dim)

L'$\mathrm{MSE}_\vartheta(T_n)$ può essere decomposto nel seguente modo

$$\mathrm{MSE}_\vartheta(T_n) = \mathrm{Var}_\vartheta(T_n) + [\mathrm{E}_\vartheta(T_n) - \kappa(\vartheta)]^2$$

Per dimostrare questa decomposizione è sufficiente interpretare l'$\mathrm{MSE}_\vartheta(T_n)$ come il momento secondo di $U = T_n - \kappa(\vartheta)$ e applicare le proprietà della varianza:

$$\begin{aligned}\mathrm{MSE}_\vartheta(T_n) &= \mathrm{E}_\vartheta(U^2) = \mathrm{Var}_\vartheta(U) + (\mathrm{E}_\vartheta(U))^2 \\ &= \mathrm{Var}_\vartheta(T_n - \kappa(\vartheta)) + [\mathrm{E}_\vartheta(T_n - \kappa(\vartheta))]^2 \\ &= \mathrm{Var}_\vartheta(T_n) + [\mathrm{E}_\vartheta(T_n) - \kappa(\vartheta)]^2\end{aligned}$$

Osservate che tutti i passaggi sono leciti perché l'MSE di uno stimatore T_n esiste se e solo se T_n ha media e varianza finite o, equivalentemente, se e solo se ha momento secondo $\mathrm{E}_\vartheta(T_n^2)$ finito.

Soluzione L'errore quadratico medio (mean square error) è definito da

$$\mathrm{MSE}_{T_n}(\theta) = \mathrm{E}_\theta\left[\left(T_n - \kappa(\theta)\right)^2\right].$$

Togliamo e aggiungiamo la media $\mathrm{E}_\theta[T_n]$ all'interno del quadrato, sviluppiamo il quadrato, usiamo la linearità del valore atteso e il fatto che la media di una costante è la costante:

$$\begin{aligned}\mathrm{MSE}_{T_n}(\theta) &= \mathrm{E}_\theta\left[\left(T_n - \mathrm{E}_\theta[T_n] + \mathrm{E}_\theta[T_n] - \kappa(\theta)\right)^2\right] \\ &= \mathrm{E}_\theta\left[\left(T_n - \mathrm{E}_\theta[T_n]\right)^2 + \left(\mathrm{E}_\theta[T_n] - \kappa(\theta)\right)^2 + 2\left(T_n - \mathrm{E}_\theta[T_n]\right)\left(\mathrm{E}_\theta[T_n] - \kappa(\theta)\right)\right] \\ &= \mathrm{E}_\theta\left[\left(T_n - \mathrm{E}_\theta[T_n]\right)^2\right] + \left(\mathrm{E}_\theta[T_n] - \kappa(\theta)\right)^2 + 2\left(\mathrm{E}_\theta[T_n] - \kappa(\theta)\right)\mathrm{E}_\theta\left[T_n - \mathrm{E}_\theta[T_n]\right].\end{aligned}$$

Essendo $\mathrm{E}_\theta\left[T_n - \mathrm{E}_\theta[T_n]\right] = 0$ ed $\mathrm{E}_\theta\left[\left(T_n - \mathrm{E}_\theta[T_n]\right)^2\right] = \mathrm{Var}_\theta[T_n]$, abbiamo ottenuto che l'errore quadratico medio è uguale alla varianza dello stimatore più la distorsione al quadrato:

$$\mathrm{MSE}_{T_n}(\theta) = \mathrm{Var}_\theta[T_n] + \left(\mathrm{E}_\theta[T_n] - \kappa(\theta)\right)^2.$$

La successione di stimatori è detta consistente in media quadratica se

$$\lim_{n\to+\infty} \mathrm{MSE}_{T_n}(\theta) = 0, \qquad \forall\theta \in \Theta.$$

Dato che l'errore quadratico può essere espresso come somma di quantità positive, entrambe queste quantità devono tendere a zero. Dunque la successione di stimatori è consistente in media quadratica se e solo se vale la non distorsione asintotica ($\lim_{n\to+\infty} \mathrm{E}_\theta[T_n] = k(\theta)$, $\forall\theta$) e la successione delle varianze tende a zero. Dalla teoria sappiamo che la consistenza in media quadratica implica anche la consistenza semplice (o debole). ▮

Intervalli esatti per campioni esponenziali negativi (con dimostrazione)

Siano X_i iid $Exp(\lambda)$, allora per le proprietà sulle somme di Gamma si ha $\sum_{i=1}^n X_i \sim Gamma(n, \lambda)$, e dunque (verificarlo)

$$Q_n = \frac{2}{\lambda} \sum_{i=1}^n X_i \sim Gamma(n, 1/2) = \chi^2_{2n}$$

(si ricordi che λ in una v.a. Gamma è un parametro di rate, ossia $1/\lambda$ è un parametro di scala). In altri termini Q_n è un possibile pivot, ossia

$$1 - \alpha = P_\lambda\{\chi_{2n,1-\alpha/2} \leq Q_n \leq \chi_{2n,\alpha/2}\}$$

Dove $\chi_{2n,\alpha}$ è il quantile (destro) di un χ^2_{2n}. Invertendo le disuguaglianze $\cdot\chi_{2n,1-\alpha/2} < Q_n < \chi_{2n,1-\alpha/2}$ si ha che

$$1 - \alpha = P_\lambda\Big\{ \frac{2}{\chi_{2n,\alpha/2}} \sum_{i=1}^n X_i \leq \lambda \leq \frac{2}{\chi_{2n,1-\alpha/2}} \sum_{i=1}^n X_i \Big\}$$

ossia un intervallo di confidenza di livello $1 - \alpha$ per λ è dato da

$$\Big[\frac{2}{\chi_{2n,\alpha/2}} \sum_{i=1}^n X_i, \frac{2}{\chi_{2n,1-\alpha/2}} \sum_{i=1}^n X_i \Big]$$