Esercizi di CP2, V
a.a. 2001/2002

Esercizio 1 Sia $X \sim \text{Un}(0,10)$ e[1] $N = [X]$.

a) Calcolare $\mathbb{P}(X \le x, N = n)$, per ogni $x \in \mathbb{R}$ e $n \in \mathbb{N}$.

b) Fissato $n \in \mathbb{N}$, si indichi con Q_n la legge condizionata di X dato il verificarsi dell'evento $\{N = n\}$:

$$Q_n(A) = \mathbb{P}(X \in A \,|\, N = n), \quad A \in \mathcal{B}(\mathbb{R}).$$

Determinare tutti i valori di n per i quali Q_n è ben posta; verificare che per tali valori Q_n è una misura di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ e scrivere esplicitamente la funzione di distribuzione condizionata:

$$F_n(x) = Q_n((-\infty, x]) = \mathbb{P}(X \le x \,|\, N = n), \qquad x \in \mathbb{R}.$$

c) Mostrare che esiste un insieme Γ tale che $\mathbb{P}(N \in \Gamma) = 1$ e tale che, per ogni $n \in \Gamma$, la legge condizionata di X dato $\{N = n\}$ ha densità (dire se si tratta di una densità nota).

Esercizio 2 Siano $p_1(x) = a\,x^\alpha\, 1\!\!1_{x>1}$ e $p_2(x) = b\,|x|^\beta\, 1\!\!1_{-1<x\le 1}$. Determinare per quali valori dei parametri $a, \alpha, b, \beta \in \mathbb{R}$ le funzioni p_1 e p_2 sono densità di probabilità di variabili aleatorie che appartengono a L^p, con $p \ge 1$ fissato.

Esercizio 3 Si supponga che tutte le v.a. coinvolte nel seguito siano in L^2. Dimostrare che:

a) $\text{Var}(X) = 0$ se e solo se X è costante q.c.;

b) $\text{Cov}(X, X) = \text{Var}(X)$;

c) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;

d) $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$, per ogni $a \in \mathbb{R}$;

e) $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.

Dedurre che, per ogni $a, b, c, d \in \mathbb{R}$,

$$\text{Cov}(aX_1 + bX_2, cY_1 + dY_2) = ac\text{Cov}(X_1, Y_1) + ad\text{Cov}(X_1, Y_2) + bc\text{Cov}(X_2, Y_1) + bd\text{Cov}(X_2, Y_2)$$

(ovvero, la covarianza è un'applicazione *bilineare*) e che

$$\text{Var}(X + c) = \text{Var}(X), \quad \text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

Esercizio 4 Siano X e Y v.a. in L^2.

a) Posto, per $\alpha \in \mathbb{R}$, $\psi(\alpha) = ||X - \alpha||_2$, dimostrare che

$$\inf_{\alpha \in \mathbb{R}} \psi(\alpha) = \min_{\alpha \in \mathbb{R}} \psi(\alpha) = \psi(\alpha^*), \qquad \text{con } \alpha^* = \mathbb{E}(X).$$

Dedurre che $\mathbb{E}(X)$ è la costante che approssima meglio in L^2 una v.a. $X \in L^2$ e che lo scarto quadratico medio (=radice della varianza) dà l'errore di approssimazione.

b) Posto, per $a, b \in \mathbb{R}$, $\varphi(x) = a\,x + b$ e $\psi(a, b) = ||Y - \varphi(X)||_2$, dimostrare che

$$\inf_{a,b \in \mathbb{R}} \psi(a, b) = \min_{a,b \in \mathbb{R}} \psi(a, b) = \psi(a^*, b^*), \qquad \text{con } a^* = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \text{ e } b^* = \mathbb{E}(Y) - a^*\mathbb{E}(X).$$

Dedurre che la tra tutte le possibili trasformazioni affini $aX + b$ di X, la v.a. $a^*X + b^*$ è quella che approssima meglio Y in L^2. La retta $y = a^*x + b^*$ prende il nome di **retta di regressione**.

Calcolare l'errore di approssimazione e dedurre infine la disuguaglianza

$$|\text{Cov}(X, Y)| \le \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}.$$

---

[1] Il simbolo $[\cdot]$ denota la *parte intera*.

Esercizio1 a) Poiché $X \in (0,10)$ q.c., $N = [X] \in E = \{0,1,\dots,9\}$. Quindi, $\mathbb{P}(X \le x, N = n) = 0$ per $x \le 0$ oppure $n \notin E$. Fissiamo ora $x > 0$ e $n \in E$:

$$\begin{aligned}\mathbb{P}(X \le x, N = n) &= \mathbb{P}(X \le x, [X] = n) = \mathbb{P}(X \le x, n \le X < n+1) \\ &= \begin{cases} 0 & \text{se } x < n; \\ \mathbb{P}(n \le X \le x) & \text{se } n \le x < n+1; \\ \mathbb{P}(n \le X < n+1) & \text{se } x \ge n+1. \end{cases}\end{aligned}$$

Ma

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0; \\ \frac{x}{10} & \text{se } 0 \le x < 10; \\ 1 & \text{se } x \ge 10, \end{cases}$$

e dunque

$$\mathbb{P}(X \le x, N = n) = \begin{cases} \frac{x-n}{10} & \text{se } n \in E \text{ e } n \le x < n+1; \\ \frac{1}{10} & \text{se } n \in E \text{ e } x \ge n+1; \\ 0 & \text{altrimenti.} \end{cases}$$

b) Q_n è ben posta se e solo se l'evento condizionante ha probabilità positiva, cioè $\mathbb{P}(N = n) > 0$, quindi per $n \in E$. Poiché inoltre Q_n rappresenta una probabilità condizionata, è essa stessa una misura di probabilità[2]. In tal caso, fissato $n \in E$,

$$F_n(x) = \mathbb{P}(X \le x | N = n) = \frac{\mathbb{P}(X \le x, N = n)}{\mathbb{P}(N = n)}$$

per $x \in \mathbb{R}$. Usando **a)** e osservando che $\mathbb{P}(N = n) = \mathbb{P}(n \le X < n+1) = 1/10$ per $n \in E$, si ha

$$F_n(x) = \begin{cases} 0 & \text{se } x < n; \\ x - n & \text{se } n \le x < n+1; \\ 1 & \text{se } x \ge n+1. \end{cases}$$

c) Basta prendere $\Gamma = E$: in tal caso $F_n(x) = \int_{-\infty}^{x} \mathbb{1}_{(n,n+1)}(t)dt$, quindi Q_n ha densità $f_n(x) = \mathbb{1}_{(n,n+1)}(x)$, che corrisponde alla densità di una v.a. $\mathrm{Un}(n, n+1)$. In termini probabilistici, ciò significa che condizionatamente all'evento $\{N = n\}$ (ovvero: noto che la parte intera di X è n), la v.a. X è uniforme su $(n, n+1)$, come d'altro canto suggerirebbe l'intuizione.

Esercizio2 Dev'essere $p_i \ge 0$ q.o., quindi $a, b \ge 0$, e

$$\int_{\mathbb{R}} p_i(x)dx = 1 \qquad \int_{\mathbb{R}} |x|^p p_i(x)dx < \infty.$$

[2] Infatti:

- $Q_n(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R} | N = n) = 1$;
- $Q_n(A) \ge 0$, per ogni $A \in \mathcal{B}(\mathbb{R})$;
- se $\{A_k\}_k \subset \mathcal{B}(\mathbb{R})$ è tale che $A_{k_1} \cap A_{k_2}$ per ogni $k_1 \ne k_2$, allora

$$\begin{aligned} Q_n(\cup_k A_k) = \mathbb{P}(X \in \cup_k A_k | N = n) &= \frac{\mathbb{P}(X \in \cup_k A_k, N = n)}{\mathbb{P}(N = n)} = \frac{\sum_k \mathbb{P}(X \in A_k, N = n)}{\mathbb{P}(N = n)} \\ &= \sum_k \mathbb{P}(X \in A_k | N = n) = \sum_k Q_n(A_k) \end{aligned}$$

dunque Q_n è una misura di probabilità su $\mathcal{B}(\mathbb{R})$.

Ora, $p_1 \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, \text{Leb})$ se e solo se $\alpha + 1 < 0$ e in tal caso $\int_{\mathbb{R}} p_1(x)dx = a\int_1^\infty x^\alpha dx = -a/(\alpha+1)$, da cui segue che $a = -(\alpha+1)$. Infatti, indicando con $p_{1n}(x) = p_1(x)1\!\!1_{x\in(0,n)}$, allora $0 \le p_{1n}(x) \uparrow p_1(x)$ per $n \to \infty$, quindi (da MON) $\int_{\mathbb{R}} p_{1n}(x)dx \uparrow \int_{\mathbb{R}} p_1(x)dx$ per $n \to \infty$. Ma

$$\int_{\mathbb{R}} p_{1n}(x)dx = \begin{cases} a\, \frac{x^{\alpha+1}}{\alpha+1}\Big|_1^n = \frac{a}{\alpha+1}(n^{\alpha+1}-1) & \text{se } \alpha \neq -1 \\ a\, \frac{\ln x}{\alpha+1}\Big|_1^n = \frac{a}{\alpha+1}\ln n & \text{se } \alpha = -1 \end{cases}$$

Allora $\int_{\mathbb{R}} p_1(x)dx < \infty$ se e solo se $\alpha < -1$ e in tal caso $\int_{\mathbb{R}} p_1(x)dx = -a/(\alpha+1)$.

Analogamente, $|x|^p p_1(x) = ax^{\alpha+p}1\!\!1_{x>1} \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, \text{Leb})$ se e solo se $\alpha + p < -1$, quindi dev'essere $\alpha < -1-p$ e in tal caso $\mathbb{E}(|X|^p) = (\alpha+1)/(\alpha+p+1)$.

Discorso simile per p_2: posto $p_{2n}(x) = p_2(x)1\!\!1_{1/n<|x|<1}$, allora $0 \le p_{2n}(x) \uparrow p_2(x)$ q.o. per $n \to \infty$ e (ancora MON) $\int_{\mathbb{R}} p_{2n}(x)dx \uparrow \int_{\mathbb{R}} p_2(x)dx$. Ora

$$\int_{\mathbb{R}} p_{2n}(x)dx = 2\int_{1/n}^1 x^\beta dx = \begin{cases} 2b\, \frac{x^{\beta+1}}{\beta+1}\Big|_{1/n}^1 = \frac{2b}{\beta+1}(1-1/n^{\beta+1}) & \text{se } \beta \neq -1 \\ 2b\, \frac{\ln x}{\beta+1}\Big|_{1/n}^1 = \frac{2b}{\beta+1}\ln n & \text{se } \beta = -1 \end{cases}$$

Quindi, p_2 è integrabile se e solo se $\beta > -1$ e in tal caso $\int_{\mathbb{R}} p_2(x)dx = 2b/(\beta+1)$, da cui segue che $b = (\beta+1)/2$. Inoltre, perché anche $|x|^p p_1(x)$ sia integrabile, dev'essere anche $p + \beta > -1$, il che è vero per ogni $\beta > -1$. Si noti infatti che, detta X una v.a. con densità p_2, allora $|X| \le 1$ q.c., cioè X è limitata, quindi anche $|X|^p$ lo è e dunque ha media, cioè $|x|^p p_2(x)$ è integrabile su $\mathbb{R}$ rispetto a Leb.

Esercizio 3. a) Se $\text{Var}(X) = 0$ allora $\mathbb{E}[(X-\mathbb{E}[X])^2] = 0$, ma $(X-\mathbb{E}[X])^2 \ge 0$ quindi $(X-\mathbb{E}[X])^2 = 0$, cioè $X = \mathbb{E}[X] = cost$. Viceversa, se $X = c$ q.c. allora $\mathbb{E}[X] = c$, quindi $\text{Var}(X) = \mathbb{E}[(X-\mathbb{E}[X])^2] = 0$.

b) $\text{Cov}(X,X) = \mathbb{E}[X \cdot X] - \mathbb{E}[X]\mathbb{E}[X] = \text{Var}(X)$.

c) $\text{Cov}(X,Y) = \mathbb{E}[X\,Y] - \mathbb{E}[X]\,\mathbb{E}[Y] = \text{Cov}(Y,X)$.

d) $\text{Cov}(aX,Y) = \mathbb{E}[aX\,Y] - \mathbb{E}[aX]\,\mathbb{E}[Y] = a\mathbb{E}[X\,Y] - a\mathbb{E}[X]\,\mathbb{E}[Y] = a\text{Cov}(X,Y)$.

e) $\text{Cov}(X_1+X_2,Y) = \mathbb{E}[(X_1+X_2)Y] - \mathbb{E}[(X_1+X_2)]\,\mathbb{E}[Y] = \mathbb{E}[X_1\,Y] + \mathbb{E}[X_2\,Y] - \mathbb{E}[X_1]\,\mathbb{E}[Y] + \mathbb{E}[X_2]\,\mathbb{E}[Y] = \text{Cov}(X_1,Y) + \text{Cov}(X_2,Y)$.

Segue quindi immediatamente che la covarianza è un'applicazione simmetrica e bilineare. Applicando tale proprietà,

$$\text{Var}(X+c) = \text{Var}(X) + 2\text{Cov}(X,c) + \text{Var}(c) = \text{Var}(X)$$

perché ogni v.a. è indipendente dalle costanti. Inoltre, poiché $\text{Var}(aX+bY) = \text{Cov}(aX+bY, aX+bY)$ anche l'ultima proprietà segue dalla bilinearità della covarianza.

Esercizio 4. a) Lavoriamo con $g = \psi^2$:

$$g(\alpha) = \mathbb{E}[X^2] + \alpha^2 - 2\alpha\mathbb{E}[X],$$

quindi $g'(\alpha) = 2(\alpha - \mathbb{E}[X]) \ge 0$ se e solo se $\alpha \ge \mathbb{E}(X)$. Quindi $\alpha^* = \mathbb{E}(X)$ è il punto di minimo assoluto, cioè $\mathbb{E}(X)$ è la costante che approssima meglio la v.a. X in L^2 e l'errore di approssimazione è

$$\psi(\alpha^*) = \Big(\mathbb{E}((X-\mathbb{E}(X))^2)\Big)^{1/2} = \sqrt{\text{Var}(X)}.$$

In particolare, $\text{Var}(X)$ dà un'indicazione di quanto la v.a. X è "dispersa" intorno alla sua media.

b) Anche qui lavoriamo con $g = \psi^2$. Sviluppando il quadrato, si ottiene facilmente che

$$g(a,b) = \mathbb{E}[Y^2] + a^2\mathbb{E}[X^2] + b^2 - 2a\mathbb{E}[XY] - 2b\mathbb{E}[Y] + 2ab\mathbb{E}[X],$$

quindi $\nabla g = 2(a\mathbb{E}[X^2] - \mathbb{E}[XY] + b\mathbb{E}[X], b - \mathbb{E}[Y] + a\mathbb{E}[X])$. L'equazione $\nabla g = 0$ ha come unica soluzione

$$a^* = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}, \qquad b^* = \mathbb{E}(Y) - a^*\mathbb{E}(X).$$

La matrice hessiana

$$H = 2\begin{pmatrix} \mathbb{E}[X^2] & \mathbb{E}[X] \\ \mathbb{E}[X] & 1 \end{pmatrix}$$

è definita positiva[3], quindi (a^*, b^*) è un punto di minimo relativo per g. Osservando infine che g è quadratica nelle variabili e che tende a $+\infty$ quando $(a,b) \notin [-n,n] \times [-n,n]$ e $n \to +\infty$, (a^*, b^*) è in effetti un punto di minimo assoluto.

L'errore di approssimazione è dato da $\psi(a^*, b^*)$:

$$\psi(a^*, b^*) = \Big(\mathbb{E}[(Y - a^*X - b^*)^2]\Big)^{1/2} = \Big(\mathbb{E}[((Y - \mathbb{E}(Y)) - a^*(X - \mathbb{E}(X)))^2]\Big)^{1/2}$$
$$= \Big(\mathrm{Var}(Y) + a^{*\,2}\mathrm{Var}(X) - 2a^*\mathrm{Cov}(X,Y)\Big)^{1/2} = \Big(\frac{\mathrm{Var}(X)\mathrm{Var}(Y) - \mathrm{Cov}^2(X,Y)}{\mathrm{Var}(X)}\Big)^{1/2}.$$

Inoltre, poiché $\psi^2 \geq 0$, dev'essere $\mathrm{Var}(X)\mathrm{Var}(Y) - \mathrm{Cov}^2(X,Y) \geq 0$, da cui segue[4] che

$$|\mathrm{Cov}(X,Y)| \leq \sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}.$$

Si rimanda alla nota in coda sul significato e le conseguenze della retta di regressione.

Riportiamo qui di seguito una tabella con alcune distribuzioni notevoli. Verificare per esercizio l'esistenza della media e della varianza, date dalle espressioni riportate in tabella.

Alcune distribuzioni notevoli

Nome	**Densità**	**Media**	**Varianza**
Uniforme, $\mathrm{Un}(a,b)$ $(-\infty < a < b < +\infty)$	$p(x) = \frac{1}{b-a}1\!\!1_{(a,b)}(x)$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Gaussiana, $\mathrm{N}(m, \sigma^2)$ $(m \in \mathbb{R}, \sigma^2 > 0)$	$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2
Esponenziale, $\mathrm{Exp}(\lambda)$ $(\lambda > 0)$	$p(x) = \lambda e^{-\lambda x}1\!\!1_{x>0}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma, $\Gamma(\alpha, \lambda)$ $(\alpha, \lambda > 0)$	$p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}1\!\!1_{x>0}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Beta, $\mathrm{Beta}(\alpha, \beta)$ $(\alpha > 0, \beta > 0)$	$p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}x^{\beta-1}1\!\!1_{0<x<1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta-1)}$

(si noti che $\mathrm{Exp}(\lambda) \equiv \Gamma(1, \lambda)$) dove la funzione $\Gamma(\alpha)$, definita per $\alpha > 0$, è data da

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1}e^{-t}dt$$

---

[3] Si ha

$$\xi^2\mathbb{E}[X^2] + 2\xi\eta\mathbb{E}[X] + \eta^2 = \mathbb{E}[X^2]\Big(\xi + \eta\frac{\mathbb{E}[X]}{\mathbb{E}[X^2]}\Big)^2 + \frac{\eta^2}{\mathbb{E}[X^2]}\Big(\mathbb{E}[X^2] - \mathbb{E}^2[X]\Big) > 0$$

per ogni $(\xi, \eta) \neq 0$, a meno che $\mathrm{Var}(X) = 0$, che non è il caso in questione.

[4] Tale disuguaglianza è anche conseguenza della disuguaglianza di Cauchy-Schwarz:

$$|\mathrm{Cov}(X,Y)| \leq \mathbb{E}(|X - \mathbb{E}(X)||Y - \mathbb{E}(Y)|) \leq \Big(\mathbb{E}(|X - \mathbb{E}(X)|^2)\Big)^{1/2}\Big(\mathbb{E}(|Y - \mathbb{E}(Y)|^2)\Big)^{1/2} = \sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}.$$

Nota 1 (Brevi cenni sulla retta di regressione)

Siano X e Y due v.a. sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Si vuole trovare la migliore approssimazione possibile della Y tramite una funzione lineare della X. In altre parole, si cercano due numeri $a, b \in \mathbb{R}$ tali che $aX + b$ approssima Y nel miglior modo possibile. Un problema di questo tipo appare, ad esempio, quando si vuole stimare una quantità Y che non è osservabile direttamente. Si consideri, a titolo di esempio, il caso in cui Y è un segnale e X è un'osservazione "disturbata" (vale a dire, X è uguale a Y più un errore dovuto all'osservazione).

Il primo problema consiste nel definire un criterio di ottimalità: che significato si vuole dare a "migliore approssimazione possibile"? Ovviamente, occorre partire dalla distanza tra $aX + b$ e Y, ovvero da $|aX + b - Y|$. Poichè però lavorare con i moduli può essere oneroso in termini di calcoli, si preferisce considerare $(aX+b-Y)^2$. Ora, tale quantità è aleatoria e il problema che è stato posto è "deterministico": i numeri a e b che si cercano *non* si vogliono aleatori. Per tale ragione, si considera il valore atteso di $(aX + b - Y)^2$. Dunque, il criterio di ottimalità scelto è il seguente: si vogliono determinare due numeri a e b tali da minimizzare[5]

$$(a, b) \mapsto \mathbb{E}\Big[(aX + b - Y)^2\Big] \equiv g(a, b).$$

L'esercizio mostra dunque che tale minimo è assunto in (a^*, b^*), dove

$$a^* = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} \text{ e } b^* = \mathbb{E}[Y] - a^*\mathbb{E}[X]$$

La retta di equazione $y = a^*x + b^*$, ovvero

$$y = \varphi(x) \equiv \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}(x - \mathbb{E}[X]) + \mathbb{E}[Y],$$

è detta **retta di regressione**. Dunque, in generale non è detto che $Y = \varphi(X)$ ma ciò che è vero è che la funzione φ è la funzione lineare che minimizza la distanza (vista sopra) tra Y e $\varphi(X)$.

Si noti che la retta di regressione non è "simmetrica", ovvero scambiando ruolo a X e Y cambia anche la retta di regressione. Inoltre, il coefficiente angolare della retta di regressione ha lo stesso segno della covarianza (perché $\mathrm{Var}(X)$ è sempre positiva). Ciò indica che, se la covarianza è positiva, Y tende ad assumere valori grandi in corrispondenza di valori grandi della X. Tale proprietà è detta di *dipendenza positiva*. Al contrario, se la covarianza è negativa, Y tende ad assumere valori piccoli in corrispondenza di valori grandi della X e in tal caso si dice che c'è *dipendenza negativa* tra le due v.a. Ad esempio, se si volesse creare un modello che leghi la relazione tra il peso X e l'altezza Y di un individuo preso a caso da una popolazione, ci si aspetta una covarianza positiva tra queste due v.a. perché individui alti hanno la tendenza ad essere più pesanti (ciò non toglie che possano esistere individui alti e leggeri oppure bassi e molto pesanti, eventi che, in generale, hanno probabilità di verificarsi più bassa). Ci si attende una covarianza negativa quando si considerino quantità aleatorie che hanno un effetto "antitetico" l'una rispetto all'altra.

Infine, se la covarianza è nulla (come avviene quando le due v.a. sono indipendenti), la retta di correlazione è costante, ovvero l'unica funzione lineare φ che minimizza la distanza tra Y e $\varphi(X)$ *non* dipende da X. Siamo quindi in presenza di una "forma debole" di indipendenza, e in tal caso le v.a. X e Y sono dette *non correlate* (ma attenzione: non è detto che siano indipendenti nel senso classico!).

Abbiamo appena visto come la covarianza possa essere utilizzata per misurare la dipendenza di due v.a. In realtà ciò non è del tutto vero perché la covarianza presenta un difetto:

$$\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y)$$

Ciò significa che la covarianza tra due v.a. è strettamente legata all'unità di misura: la covarianza cambia se si cambia l'unità di misura con la quale le v.a. X e Y sono state misurate. Per tale ragione, si introduce la seguente quantità:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

[5] Va detto che esistono altri criteri di ottimizzazione, che qui non saranno presi in considerazione. Si noti che ciò che si vuole minimizzare è la distanza in L^2 tra la v.a. Y e la v.a. $aX + b$.

$\rho_{X,Y}$ è detto **coefficiente di correlazione di X e Y** e *non* dipende dai cambiamenti di misurazione, cioè (esercizio!)

$$\rho_{aX,bY} = \rho_{X,Y}$$

per ogni[6] a, b tali che $ab > 0$. Inoltre, abbiamo visto che $\mathrm{Cov}(X,Y)^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$, quindi $-1 \leq \rho_{X,Y} \leq 1$.

Riassumendo, la condizione di non correlazione è in realtà molto più debole di quella di indipendenza, ma è anche più facile da verificare per cui è abbastanza usata nella pratica come forma debole di indipendenza.

[6]Se $ab < 0$, il coefficiente di correlazione cambia segno: esercizio! Tale proprietà è intuitiva se si ricorda la relazione che intercorre tra il segno del coefficiente di correlazione e la proprietà di dipendenza positiva o negativa tra le due v.a. in questione.