

ESERCITAZIONE V
 COMPLEMENTI DI PROBABILITÀ
 A.A. 2010/2011

Argomenti: spazi L^p ; misure e v.a. assolutamente continue.

Esercizio 1. Sia $f \in \mathcal{L}^1(S, \Sigma, \mu)$ tale che $\int_A f d\mu = 0$ per ogni $A \in \Sigma$. Dimostrare che $f = 0$ q.o. (i.e. $\mu(\{f \neq 0\}) = 0$).

Esercizio 2. Siano ρ, μ, ν misure σ -finite su (S, Σ) .

a) Dimostrare che se $\nu \ll \mu$ e $\mu \ll \rho$ allora $\nu \ll \rho$ e

$$\frac{d\nu}{d\rho} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\rho} \quad \rho\text{-q.o.}$$

b) Mostrare che se $\nu \equiv \mu$ allora $\frac{d\nu}{d\mu} = (\frac{d\mu}{d\nu})^{-1}$ q.o.

c) Supponiamo che $\mu \ll \rho$ e $\nu \ll \rho$ e sia $A = \{\frac{d\mu}{d\rho} = 0\} \cap \{\frac{d\nu}{d\rho} > 0\}$. Dimostrare che $\nu \ll \mu$ se e solo se $\rho(A) = 0$ e in tal caso

$$\frac{d\nu}{d\mu} = \frac{\frac{d\nu}{d\rho}}{\frac{d\mu}{d\rho}} \mathbf{1}_{\{\frac{d\mu}{d\rho} > 0\}} = \frac{\frac{d\nu}{d\rho}}{\frac{d\mu}{d\rho}} \mathbf{1}_{A^c} \quad \mu\text{-q.o.}$$

Esercizio 3. Siano μ_1, μ_2 due misure di probabilità (ad esempio) su $(\mathbb{R}, \mathcal{B})$, $\alpha \in (0, 1)$ e $\mu = \alpha \mu_1 + (1 - \alpha) \mu_2$. Mostrare che μ è ancora una misura di probabilità. Determinare inoltre delle condizioni su μ_1, μ_2 affinché

a) $\mu \ll \text{Leb}$;

b) μ è discreta;

c) μ non è né discreta né assolutamente continua rispetto a Leb.

Usando c), mostrare che esistono variabili aleatorie che pur non assumendo un numero discreto di valori, non hanno densità di probabilità.

Esercizio 4. Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e $A_0 \in \mathcal{F}$ tale che $\mathbb{P}(A_0) > 0$. Sia

$$\mathcal{F} \ni A \mapsto \mathbb{P}_0(A) = \mathbb{P}(A | A_0) = \frac{\mathbb{P}(A \cap A_0)}{\mathbb{P}(A_0)}.$$

a) Verificare che $(\Omega, \mathcal{F}, \mathbb{P}_0)$ è uno spazio di probabilità e che $\mathbb{P}_0 \ll \mathbb{P}$. Scrivere $\frac{d\mathbb{P}_0}{d\mathbb{P}}$.

b) Mostrare che $\mathbb{P} \ll \mathbb{P}_0$ se e solo se $\mathbb{P}(A_0) = 1$ e scrivere in tal caso $\frac{d\mathbb{P}}{d\mathbb{P}_0}$.

- c) Per $p \geq 1$, mostrare che se $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ allora $X \in L^p(\Omega, \mathcal{F}, \mathbb{P}_0)$.
 Detta $\mathbb{E}_0$ la media rispetto a $\mathbb{P}_0$, dimostrare che esiste una costante universale (cioè, indipendente da p) $c > 0$ tale che $\mathbb{E}_0(|X|^p) \leq c \mathbb{E}(|X|^p)$.
- d) Posto invece,

$$\mathcal{F} \ni A \mapsto \mathbb{P}^0(A) = \begin{cases} \mathbb{P}(A_0 | A) & \text{se } \mathbb{P}(A) > 0 \\ 0 & \text{se } \mathbb{P}(A) = 0 \end{cases}$$

dire se $\mathbb{P}^0$ è una misura, eventualmente di probabilità, su $(\Omega, \mathcal{F})$.

Esercizio 5. Siano $\Omega = \{0, 1, 2, \dots\}$ e $\mathcal{F} = \mathcal{P}(\Omega)$. Su $(\Omega, \mathcal{F})$, siano $\mathbb{P}_1$ e $\mathbb{P}_2$ definite da

$$\mathbb{P}_i(A) = \sum_{k \in A} p_i(k), \quad A \subset \Omega, \quad i = 1, 2$$

(con $\mathbb{P}_i(\emptyset) := 0$) dove, per $\rho \in (0, 1)$ e $N \in \mathbb{N}$ fissati,

$$(\text{misura geometrica}) \quad p_1(k) = \rho^k (1 - \rho), \quad k = 0, 1, \dots$$

$$(\text{misura binomiale}) \quad p_2(k) = \begin{cases} \binom{N}{k} \rho^k (1 - \rho)^{N-k} & k = 0, \dots, N \\ 0 & k > N \end{cases}$$

- a) Verificare che $\mathbb{P}_1$ e $\mathbb{P}_2$ sono due misure di probabilità.

- b) Mostrare che $\mathbb{P}_1 \not\ll \mathbb{P}_2$ ma che $\mathbb{P}_2 \ll \mathbb{P}_1$, e scrivere $\frac{d\mathbb{P}_2}{d\mathbb{P}_1}$.

Esercizio 6. Siano $\Omega = \{0, 1, 2, \dots\}$ e $\mathcal{F} = \mathcal{P}(\Omega)$. Su $(\Omega, \mathcal{F})$, siano $\mathbb{P}$ e $\mathbb{Q}$ le misure di probabilità rispettivamente definite da

$$\mathbb{P} = \sum_i p_i \delta_{\{i\}} \quad \text{e} \quad \mathbb{Q} = \sum_i q_i \delta_{\{i\}}$$

con $\{p_i\}_i$ e $\{q_i\}_i$ tali che $p_i, q_i \geq 0$ per ogni i e $1 = \sum_i p_i = \sum_i q_i$. Definiamo

$$\Gamma_{\mathbb{P}} = \{i : p_i > 0\} = \{i : p_i = 0\}^c = \mathcal{N}_{\mathbb{P}}^c$$

$$\Gamma_{\mathbb{Q}} = \{i : q_i > 0\} = \{i : q_i = 0\}^c = \mathcal{N}_{\mathbb{Q}}^c$$

- a) Dimostrare che $\mathbb{Q} \ll \mathbb{P}$ se e solo se $\Gamma_{\mathbb{P}} \supset \Gamma_{\mathbb{Q}}$ se e solo se $\mathcal{N}_{\mathbb{P}} \subset \mathcal{N}_{\mathbb{Q}}$.
 Nell'ipotesi $\mathbb{Q} \ll \mathbb{P}$, scrivere $\frac{d\mathbb{Q}}{d\mathbb{P}}$.

- b) Dimostrare che $\mathbb{Q} \equiv \mathbb{P}$ se e solo se $\Gamma_{\mathbb{P}} = \Gamma_{\mathbb{Q}}$ se e solo se $\mathcal{N}_{\mathbb{P}} = \mathcal{N}_{\mathbb{Q}}$.

Esercizio 7. Sia $G : \mathbb{R} \rightarrow \mathbb{R}$ una funzione monotona non decrescente, limitata e càdlàg¹, e sia

$$\mu_G((a, b]) := G(b) - G(a), \quad a, b \in \mathbb{R}, \quad a \leq b.$$

¹càdlàg=continua a destra, limitata a sinistra, ovvero $G(x^+) = G(x)$ e $G(x^-)$ esiste finito, per ogni x .

- a) Mostrare che esiste un'unica misura μ su $(\mathbb{R}, \mathcal{B})$ tale che $\mu \equiv \mu_G$ sugli intervalli della forma $(a, b]$.
- b) Trovare condizioni su G affinché $(\mathbb{R}, \mathcal{B}, \mu)$ sia uno spazio di probabilità. Se μ è una misura di probabilità, è vero che esiste una v.a. X con legge μ ? E in caso affermativo, che relazione c'è tra G e la funzione di distribuzione F_X di X ?
- c) Supponiamo $G(x) = 1 + \int_{-\infty}^x g(t) dt$, con $g \in L^1(\mathbb{R}, \mathcal{B}, \text{Leb})$ fissata e $g(t) \geq 0$ per q.o. t . Verificare che G soddisfa le ipotesi richieste e che in tal caso

$$\mu(A) = \int_A g(t) dt, \quad \text{per ogni } A \in \mathcal{B}.$$

Dedurre quindi che $\mu \ll \text{Leb}$ e scrivere $\frac{d\mu}{d\text{Leb}}$. Se μ è una misura di probabilità ed è la legge di una v.a. X , è vero che X ha densità? Se sì, quanto vale la densità f_X di X ?

Esercizio 8. Siano X e Y v.a. indipendenti. Dimostrare che se $X, Y \in \mathcal{L}^1$ allora $XY \in \mathcal{L}^1$ e $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

[Sugg: Supporre dapprima che X e Y siano non negative ed assumano un numero finito di valori. Usando MON, dimostrare l'asserto quando $X, Y \geq 0$ e passare poi al caso generale ponendo $X = X^+ - X^-$ e $Y = Y^+ - Y^-$.]

Esercizio 9. Siano $\Omega = \mathbb{R}$ e $\mathcal{F} = \mathcal{B}(\mathbb{R})$. Su $(\Omega, \mathcal{F})$, siano $\mathbb{P}_1$ e $\mathbb{P}_2$ definite da

$$\mathbb{P}_i(A) = \int_A p_i(x) dx, \quad A \subset \Omega, \quad i = 1, 2$$

dove, fissato $\lambda > 0$,

$$\begin{aligned} \text{(misura esponenziale)} \quad p_1(x) &= \lambda e^{-\lambda x} \mathbf{1}_{x>0} \\ \text{(misura uniforme)} \quad p_2(x) &= \frac{1}{\lambda} \mathbf{1}_{0<x<\lambda} \end{aligned}$$

- a) Verificare che $\mathbb{P}_1$ e $\mathbb{P}_2$ sono due misure di probabilità.
- b) Dire se $\mathbb{P}_i \ll \mathbb{P}_j$ ed in caso affermativo scrivere $\frac{d\mathbb{P}_i}{d\mathbb{P}_j}$, $i \neq j$, $i, j \in \{1, 2\}$.

Esercizio 10. Su $(\Omega, \mathcal{F}, \mathbb{P})$, sia X una v.a. di legge $\text{Exp}(2)$. Sia $\mathbb{Q}$ la misura su $(\Omega, \mathcal{F})$ definita da

$$\mathbb{Q}(A) = \mathbb{E}(2X \mathbf{1}_A), \quad A \in \mathcal{F}.$$

- a) Verificare che $\mathbb{Q}$ è una misura di probabilità.
- b) Dire se $\mathbb{Q} \ll \mathbb{P}$ e/o $\mathbb{P} \ll \mathbb{Q}$ e in caso affermativo scrivere $\frac{d\mathbb{Q}}{d\mathbb{P}}$ e/o $\frac{d\mathbb{P}}{d\mathbb{Q}}$.
- c) Per quali valori di p si ha $Y = e^X \in L^p(\Omega, \mathcal{F}, \mathbb{Q})$?

- d) Dimostrare che se Z è indipendente da X allora la legge di Z in $(\Omega, \mathcal{F}, \mathbb{Q})$ coincide con la legge di Z in $(\Omega, \mathcal{F}, \mathbb{P})$.

Esercizio 11. Siano $p_1(x) = a x^\alpha \mathbf{1}_{x>1}$ e $p_2(x) = b |x|^\beta \mathbf{1}_{-1 < x \leq 1}$. 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 $p \geq 1$ fissato.

Esercizio 12. Sia $X \sim \text{Un}(0, 10)$ e² $N = [X]$.

a) Calcolare $\mathbb{P}(X \leq x, N = n)$, 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:

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

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 13. Siano X e Y v.a. in L^2 .

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

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

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

$$\inf_{a, b \in \mathbb{R}} \psi(a, b) = \min_{a, b \in \mathbb{R}} \psi(a, b) = \psi(a^*, b^*), \quad \text{con}$$

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

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

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

²Il simbolo $[\cdot]$ denota la *parte intera*.

SOLUZIONI

Esercizio 1 $\{f > 0\} \in \Sigma$, quindi $0 = \int_{\{f > 0\}} f \, d\mu = \int f \mathbf{1}_{\{f > 0\}} \, d\mu$. Poiché $f \mathbf{1}_{\{f > 0\}} \in (m\Sigma)^+$, allora $\mu(\{f > 0\}) = 0$. Analogamente, $\{f < 0\} \in \Sigma$, quindi $0 = -\int_{\{f < 0\}} f \, d\mu = \int -f \mathbf{1}_{\{f < 0\}} \, d\mu$ e $-f \mathbf{1}_{\{f < 0\}} \in (m\Sigma)^+$, allora $\mu(\{f < 0\}) = 0$. Infine, $\mu(\{f \neq 0\}) = \mu(\{f > 0\} \cup \{f < 0\}) = 0$.

Esercizio 2 a) Se $\rho(A) = 0$ allora $\mu(A) = 0$ ($\mu \ll \rho$), e quindi $\nu(A) = 0$ ($\nu \ll \mu$). Ciò prova che $\nu \ll \rho$. Dobbiamo mostrare che

$$\nu(A) = \int_A \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\rho} \, d\rho, \quad A \in \Sigma.$$

Abbiamo visto che se $\mu \ll \rho$ allora $h \in \mathcal{L}^1(S, \Sigma, \mu)$ se e solo se $h \cdot \frac{d\mu}{d\rho} \in \mathcal{L}^1(S, \Sigma, \rho)$ e in tal caso

$$\int h \, d\mu = \int h \cdot \frac{d\mu}{d\rho} \, d\rho.$$

Ora, $\frac{d\nu}{d\mu} \in \mathcal{L}^1(S, \Sigma, \mu)$, quindi $\frac{d\nu}{d\mu} \mathbf{1}_A \in \mathcal{L}^1(S, \Sigma, \mu)$ per ogni $A \in \Sigma$, e allora

$$\nu(A) = \int_A \frac{d\nu}{d\mu} \, d\mu = \int \frac{d\nu}{d\mu} \mathbf{1}_A \, d\mu = \int \frac{d\nu}{d\mu} \mathbf{1}_A \cdot \frac{d\mu}{d\rho} \, d\rho = \int_A \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\rho} \, d\rho.$$

b) Se $\mu \equiv \nu$ allora ovviamente $\nu \ll \mu$ e $\mu \ll \nu$: usando **a)** con $\rho = \nu$, si ottiene

$$1 = \frac{d\nu}{d\nu} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\nu}.$$

c) Osserviamo che $\mu(A) = \int_A \frac{d\mu}{d\rho} \, d\rho \leq \int_{\{\frac{d\mu}{d\rho} = 0\}} \frac{d\mu}{d\rho} \, d\rho = 0$, quindi $\mu(A) = 0$.

Supponiamo che $\nu \ll \mu$: in particolare allora $0 = \nu(A) = \mu(A) = \int_A \frac{d\nu}{d\rho} \, d\rho$. Poiché $\frac{d\nu}{d\rho} \geq 0$ ρ -q.o., dev'essere $\frac{d\nu}{d\rho} \mathbf{1}_A = 0$ ρ -q.o. Ma

$$\left\{ \frac{d\nu}{d\rho} \mathbf{1}_A > 0 \right\} = A \cap \left\{ \frac{d\nu}{d\rho} > 0 \right\} = A,$$

quindi dev'essere $\rho(A) = 0$.

Viceversa, supponiamo che $\rho(A) = 0$. Sia $N \in \Sigma$ tale che $\mu(N) = 0$, cioè $0 = \mu(N) = \int_N \frac{d\mu}{d\rho} \, d\rho = \int_{N \cap \{\frac{d\mu}{d\rho} > 0\}} \frac{d\mu}{d\rho} \, d\rho$. Allora,

$$\rho\left(N \cap \left\{ \frac{d\mu}{d\rho} > 0 \right\}\right) = 0.$$

Proviamo che $\nu(N) = 0$:

$$\mu(N) = \int_N \frac{d\nu}{d\rho} d\rho = \int_{N \cap \{\frac{d\nu}{d\rho} > 0\}} \frac{d\nu}{d\rho} d\rho.$$

Ora,

$$\begin{aligned} N \cap \left\{ \frac{d\nu}{d\rho} > 0 \right\} &= \left(N \cap \left\{ \frac{d\nu}{d\rho} > 0 \right\} \cap \left\{ \frac{d\mu}{d\rho} = 0 \right\} \right) \cup \left(N \cap \left\{ \frac{d\nu}{d\rho} > 0 \right\} \cap \left\{ \frac{d\mu}{d\rho} > 0 \right\} \right) \\ &\subset A \cup \left(N \cap \left\{ \frac{d\mu}{d\rho} > 0 \right\} \right) \end{aligned}$$

quindi $\rho(N \cap \{\frac{d\nu}{d\rho} > 0\}) \leq \rho(A) + \rho(N \cap \{\frac{d\mu}{d\rho} > 0\}) = 0$. Ma allora

$$\mu(N) = \int_{N \cap \{\frac{d\nu}{d\rho} > 0\}} \frac{d\nu}{d\rho} d\rho = 0.$$

Infine, supponendo che $\nu \ll \mu$, usando **a)** si ha ρ -q.o.

$$\frac{d\nu}{d\rho} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\rho}$$

quindi su $\{\frac{d\mu}{d\rho} > 0\}$ si ha

$$\frac{d\nu}{d\mu} = \frac{\frac{d\nu}{d\rho}}{\frac{d\mu}{d\rho}}.$$

Ora, gli insiemi $\{\frac{d\mu}{d\rho} > 0\}^c$ e A sono di ρ -misura nulla, quindi di μ -misura nulla. Allora possiamo scrivere

$$\frac{d\nu}{d\mu} = \frac{\frac{d\nu}{d\rho}}{\frac{d\mu}{d\rho}} \mathbf{1}_{\{\frac{d\mu}{d\rho} > 0\}} = \frac{\frac{d\nu}{d\rho}}{\frac{d\mu}{d\rho}} \mathbf{1}_{A^c}$$

q.o. rispetto a μ .

Esercizio 3 $\mu(A) = \alpha\mu_1(A) + (1 - \alpha)\mu_2(A) \geq 0$ per ogni $A \in \mathcal{B}$. Inoltre, se $A_1, A_2, \dots$ sono disgiunti,

$$\begin{aligned} \mu(\cup_k A_k) &= \alpha\mu_1(\cup_k A_k) + (1 - \alpha)\mu_2(\cup_k A_k) = \alpha \sum_k \mu_1(A_k) + (1 - \alpha) \sum_k \mu_2(A_k) \\ &= \sum_k \left(\alpha\mu_1(A_k) + (1 - \alpha)\mu_2(A_k) \right) = \sum_k \mu(A_k) \end{aligned}$$

quindi μ è una misura. Infine, poiché $\mu_1(\mathbb{R}) = \mu_2(\mathbb{R}) = 1$, $\mu(\mathbb{R}) = \alpha + (1 - \alpha) = 1$, cioè μ è una misura di probabilità su $(\mathbb{R}, \mathcal{B})$.

a) Dev'essere $\text{Leb}(A) = 0 \Rightarrow \mu(A) = 0$. Ora, se $0 = \mu(A) = \alpha\mu_1(A) + (1 - \alpha)\mu_2(A)$, allora $\mu_1(A) = 0$ e $\mu_2(A) = 0$. Quindi, $\mu \ll \text{Leb}$ se e solo se $\mu_1 \ll \text{Leb}$ e $\mu_2 \ll \text{Leb}$.

b) Se μ è discreta allora esiste E discreto tale che $\mu(\{x\}) > 0$ per ogni $x \in E$ e $\mu(E^c) = 0$. Quindi

$$\alpha\mu_1(E^c) + (1 - \alpha)\mu_2(E^c) = \mu(E^c) = 0,$$

e poiché $\alpha, 1 - \alpha > 0$, dev'essere $\mu_1(E^c) = 0 = \mu_2(E^c)$. Questo significa che, detti E_1 e E_2 i supporti di μ_1 e μ_2 , allora $E_1 \subset E$ e $E_2 \subset E$, cioè E_1 e E_2 sono discreti e quindi μ_1 e μ_2 sono misure discrete.

c) Basta prendere una misura discreta, ad esempio μ_1 , e l'altra assolutamente continua rispetto a Leb, cioè $\mu_2 \ll \text{Leb}$. Indichiamo con E_1 ed E_2 il supporto³ di μ_1 e μ_2 rispettivamente. Si noti che E_1 è discreto ed E_2 non lo è. In tal caso, μ non è discreta, perché $\mu(E_2) = (1 - \alpha)\mu_2(E_2) > 0$, e μ non è assolutamente continua rispetto a Leb, perché $\text{Leb}(E_1) = 0$ ma $\mu(E_1) = \alpha\mu_1(E_1) > 0$.

Infine, sia μ la misura di probabilità definita come in **c)**. Allora (per il Teorema di rappresentazione di Skorohod), esiste uno spazio di probabilità su cui si può definire una v.a. X tale che μ è la legge di X , cioè $\Lambda_X = \mu$. Ora, Λ_X non è (come misura) discreta né assolutamente continua, quindi X non è (come v.a.) discreta né assolutamente continua.

Esercizio 4 a) Si ha: $\mathbb{P}_0$ è in effetti definita su tutto $\mathcal{F}$; $\mathbb{P}_0(\Omega) = 1$; $\mathbb{P}_0(A) \geq 0$ per ogni $A \in \mathcal{F}$; se $\{A_k\}_k \subset \mathcal{F}$ è tale che $A_{k_1} \cap A_{k_2} = \emptyset$ per ogni $k_1 \neq k_2$, allora

$$\begin{aligned} \mathbb{P}_0(\cup_k A_k) &= \mathbb{P}(\cup_k A_k \mid A_0) = \frac{\mathbb{P}(\cup_k A_k \cap A_0)}{\mathbb{P}(A_0)} = \frac{\sum_k \mathbb{P}(A_k \cap A_0)}{\mathbb{P}(A_0)} \\ &= \sum_k \mathbb{P}(A_k \mid A_0) = \sum_k \mathbb{P}_0(A_k) \end{aligned}$$

dunque $\mathbb{P}_0$ è una misura di probabilità su $(\Omega, \mathcal{F})$.

Poi, se $\mathbb{P}(A) = 0$ allora $\mathbb{P}(A \cap A_0) = 0$, quindi $\mathbb{P}_0(A) = 0$, cioè $\mathbb{P}_0 \ll \mathbb{P}$. Inoltre, per ogni $A \in \mathcal{F}$,

$$\mathbb{P}_0(A) = \frac{\mathbb{P}(A \cap A_0)}{\mathbb{P}(A_0)} = \frac{1}{\mathbb{P}(A_0)} \int_{A \cap A_0} d\mathbb{P} = \int_A \frac{\mathbf{1}_{A_0}}{\mathbb{P}(A_0)} d\mathbb{P},$$

da cui segue che $\mathbb{P}$ -q.o.

$$\frac{d\mathbb{P}_0}{d\mathbb{P}}(\omega) = \frac{1}{\mathbb{P}(A_0)} \mathbf{1}_{A_0}(\omega).$$

b) Supponiamo $\mathbb{P} \ll \mathbb{P}_0$. Poiché $\mathbb{P}_0(A_0^c) = 0$, dev'essere $\mathbb{P}(A_0^c) = 0$ cioè $\mathbb{P}(A_0) = 1$.

³Il supporto di una misura su $(S, \mathcal{B}(S))$ si definisce come il più piccolo chiuso tale che $\mu(E^c) = 0$. Se μ è una misura di probabilità, il supporto è il più piccolo chiuso tale che $\mu(E) = 1$, cioè per ogni chiuso F tale che $E \subset F$ allora $\mu(F) = 1$.

Viceversa, supponiamo che $\mathbb{P}(A_0) = 1$. Allora $\mathbb{P}(A \cap A_0) = \mathbb{P}(A)$ quindi $\mathbb{P}_0(A) = \mathbb{P}(A)$ per ogni $A \in \mathcal{F}$, e ovviamente $\mathbb{P} \ll \mathbb{P}_0$.

In particolare, $\mathbb{P} \ll \mathbb{P}_0$ se e solo se $\mathbb{P}_0 \equiv \mathbb{P}$, ma allora $\frac{d\mathbb{P}}{d\mathbb{P}_0} = 1$.

c) Abbiamo visto che $\mathbb{P}_0 \ll \mathbb{P}$. Allora,

$$Z \in L^1(\Omega, \mathcal{F}, \mathbb{P}_0) \quad \text{se e solo se} \quad Z \frac{d\mathbb{P}_0}{d\mathbb{P}} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$$

e in tal caso, $\mathbb{E}_0(Z) = \int Z d\mathbb{P}_0 = \int Z \frac{d\mathbb{P}_0}{d\mathbb{P}} d\mathbb{P}$. Ora, sia $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, cioè $Z = |X|^p \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Allora,

$$\left| Z \frac{d\mathbb{P}_0}{d\mathbb{P}} \right| = \frac{1}{\mathbb{P}(A_0)} |X|^p \mathbf{1}_{A_0} \leq c |X|^p$$

dove si è posto $c = 1/\mathbb{P}(A_0)$, quindi (per dominazione) $Z \frac{d\mathbb{P}_0}{d\mathbb{P}} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Ciò prova che $|X|^p \in L^1(\Omega, \mathcal{F}, \mathbb{P}_0)$, o meglio $X \in L^p(\Omega, \mathcal{F}, \mathbb{P}_0)$, ed inoltre

$$\mathbb{E}_0(|X|^p) = c \int |X|^p \mathbf{1}_{A_0} d\mathbb{P} \leq c \int |X|^p d\mathbb{P} = c \mathbb{E}(|X|^p).$$

d) No: in generale $\mathbb{P}^0$ non è una misura, tanto meno di probabilità. Infatti, se $\mathbb{P}(A_0) = 1$, allora $\mathbb{P}^0(A) = 1$ per ogni $A \in \mathcal{F}$ tale che $\mathbb{P}(A) > 0$, quindi $\mathbb{P}^0$ non è additiva. Se invece $\mathbb{P}(A_0) < 1$ si prenda, ad esempio, $A = A_0$ e $B = A_0^c$. Allora, $A \cap B = \emptyset$ e $A \cup B = \Omega$, quindi

$$\mathbb{P}^0(A \cup B) = \mathbb{P}(A_0 | \Omega) = \mathbb{P}(A_0) < 1.$$

Inoltre, $\mathbb{P}^0(A) = \mathbb{P}^0(A_0) = 1$ e $\mathbb{P}^0(B) = \mathbb{P}^0(A_0^c) = 0$, quindi $\mathbb{P}^0(A \cup B) < \mathbb{P}^0(A) + \mathbb{P}^0(B)$, ovvero $\mathbb{P}^0$ non è una funzione additiva, dunque non è una misura.

Esercizio 5 a) Occorre far vedere che $\sum_k p_i(k) = 1$, $i = 1, 2$, il che è immediato (serie geometrica e binomio di Newton).

b) Sia A un sottoinsieme non vuoto di $\{N+1, N+2, \dots\}$. Ovviamente $\mathbb{P}_2(A) = 0$ mentre $\mathbb{P}_1(A) > 0$, dunque $\mathbb{P}_1 \not\ll \mathbb{P}_2$.

Ora, osserviamo che $\mathbb{P}_1(A) = 0$ se e solo se $A = \emptyset$, da cui segue che $\mathbb{P}_2 \ll \mathbb{P}_1$. Inoltre, per ogni $A \subset \Omega$, $A \neq \emptyset$,

$$\mathbb{P}_2(A) = \sum_{k \in A} p_2(k) = \sum_{k \in A} \frac{p_2(k)}{p_1(k)} p_1(k).$$

Posto quindi $g(k) = \frac{p_2(k)}{p_1(k)}$ per $k \in \Omega$, si ha

$$\mathbb{P}_2(A) = \sum_{k \in A} g(k) p_1(k) = \int_A g(\omega) \mathbb{P}_1(d\omega),$$

cioè $g = \frac{d\mathbb{P}_2}{d\mathbb{P}_1}$.

Esercizio 6 a) $\mathbb{Q} \ll \mathbb{P}$ se e solo se $\{A \in \mathcal{F} : \mathbb{P}(A) = 0\} \subset \{A \in \mathcal{F} : \mathbb{Q}(A) = 0\}$. Ora, è facile vedere che $\{A \in \mathcal{F} : \mathbb{P}(A) = 0\} = \mathcal{P}(\mathcal{N}_{\mathbb{P}})$ e $\{A \in \mathcal{F} : \mathbb{Q}(A) = 0\} = \mathcal{P}(\mathcal{N}_{\mathbb{Q}})$, quindi $\mathbb{Q} \ll \mathbb{P}$ se e solo se $\mathcal{N}_{\mathbb{P}} \subset \mathcal{N}_{\mathbb{Q}}$ o equivalentemente $\Gamma_{\mathbb{P}} \supset \Gamma_{\mathbb{Q}}$. Assumendo $\mathbb{Q} \ll \mathbb{P}$, si ha, per $A \in \mathcal{F}$,

$$\mathbb{Q}(A) = \int_A d\mathbb{Q} = \sum_{i \in A} q_i = \sum_{i \in A \cap \Gamma_{\mathbb{Q}}} q_i = \sum_{i \in A \cap \Gamma_{\mathbb{P}}} \frac{q_i}{p_i} p_i$$

Posto X tale che $X(i) = q_i/p_i$ se $i \in \Gamma_{\mathbb{P}}$ e $X(i) = 0$ altrimenti, si ha

$$\mathbb{Q}(A) = \int_A d\mathbb{Q} = \sum_{i \in A} X(i) p_i = \int_A X d\mathbb{P}$$

da cui segue che $\frac{d\mathbb{Q}}{d\mathbb{P}} = X$.

b) $\mathbb{Q} \ll \mathbb{P}$ se e solo se $\mathcal{N}_{\mathbb{P}} \subset \mathcal{N}_{\mathbb{Q}}$. $\mathbb{P} \ll \mathbb{Q}$ se e solo se $\mathcal{N}_{\mathbb{Q}} \subset \mathcal{N}_{\mathbb{P}}$. Ma allora, $\mathbb{Q} \equiv \mathbb{P}$ se e solo se $\mathcal{N}_{\mathbb{Q}} = \mathcal{N}_{\mathbb{P}}$.

Esercizio 7 a) Si tratta della stessa costruzione che consente di passare dalla f.d. alla legge. I passi sono i seguenti.

• Sia $\mathcal{A} = \{\text{unioni finite di intervalli disgiunti}\}$ e definiamo μ_G su $\mathcal{A}$:

$$\begin{aligned}\mu_G((a, b]) &= G(b) - G(a) \\ \mu_G((a, b)) &= G(b^-) - G(a) \\ \mu_G([a, b)) &= G(b^-) - G(a^-) \\ \mu_G([a, b]) &= G(b) - G(a^-)\end{aligned}$$

e se $A = \cup_{i=1}^n I_i$, con I_i intervalli disgiunti,

$$\mu_G(A) = \sum_{i=1}^n \mu_G(I_i).$$

• $\mathcal{A}$ è un'algebra e μ_G è σ -additiva e finita su $\mathcal{A}$: per il teorema di Carathéodory esiste un'unica misura μ su $\sigma(\mathcal{A}) = \mathcal{B}$ tale che $\mu \equiv \mu_G$ su $\mathcal{A}$.

b) Dev'essere $\mu(\mathbb{R}) = 1$:

$$\begin{aligned}1 = \mu(\mathbb{R}) &= \lim_{x \rightarrow +\infty} \mu((-x, x]) = \lim_{x \rightarrow +\infty} \mu_G((-x, x]) \\ &= \lim_{x \rightarrow +\infty} (G(x) - G(-x)) = G(+\infty) - G(-\infty)\end{aligned}$$

quindi il salto "globale" di G dev'essere 1. Dal teorema di rappresentazione di Skorohod, esiste almeno una v.a. X tale che $\mu \equiv \Lambda_X$. Inoltre,

$$F_X(x) = \Lambda_X((-\infty, x]) = \mu((-\infty, x]) = G(x) - G(-\infty)$$

cioè, posto $c = G(-\infty)$, dev'essere $F_X(x) = G(x) - c$, ovvero F_X è un'opportuna traslazione di G .

c) Poiché $g \geq 0$ q.o., $G(x) = 1 + \int_{-\infty}^x g(t) dt$ è una funzione monotona crescente, continua e, dal momento che g è integrabile, anche limitata, dunque verifica le ipotesi richieste. Sia

$$\nu(A) = \int_A g(t) dt, \quad A \in \mathcal{B}.$$

Ora, $\pi = \{(-\infty, x]; x \in \mathbb{R}\}$ è un π -system che genera $\mathcal{B}$ e su π

$$\begin{aligned} \mu((-\infty, x]) &= \mu_G((-\infty, x]) = G(x) - G(-\infty) \\ &= 1 + \int_{-\infty}^x g(t) dt - 1 = \int_{(-\infty, x]} g(t) dt = \nu((-\infty, x]) \end{aligned}$$

Dunque, le misure finite μ e ν coincidono sul π -system π , quindi coincidono su $\sigma(\pi) = \mathcal{B}$:

$$\mu(A) = \nu(A) = \int_A g(t) dt, \quad A \in \mathcal{B}.$$

Allora, $\mu \ll \text{Leb}$ e $\frac{d\mu}{d\text{Leb}} = g$. Se $\mu = \Lambda_X$, si ha quindi $\frac{d\Lambda_X}{d\text{Leb}} = g$, che per definizione è la (una) densità di X .

Esercizio 9 a) Poiché $p_i \geq 0$, $\mathbb{P}_i$ è una misura su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $i = 1, 2$. Poi, si ha

$$\begin{aligned} \int_{\mathbb{R}} p_1(x) dx &= \int_0^{\infty} \lambda e^{\lambda x} dx = 1 \\ \int_{\mathbb{R}} p_2(x) dx &= \int_0^{\lambda} \frac{1}{\lambda} dx = 1, \end{aligned}$$

quindi $\mathbb{P}_1$ e $\mathbb{P}_2$ sono entrambe misure di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

b) $\mathbb{P}_1 \not\ll \mathbb{P}_2$: preso, ad esempio, $A = (\lambda, \lambda + 1)$, allora $\mathbb{P}_2(A) = 0$ ma $\mathbb{P}_1(A) > 0$. Invece $\mathbb{P}_2 \ll \mathbb{P}_1$. Infatti, sia $A \in \mathcal{B}(\mathbb{R})$ tale che $\mathbb{P}_1(A) = 0$. Allora, o $\text{Leb}(A) = 0$, e quindi anche $\mathbb{P}_2(A) = 0$, oppure $\text{Leb}(A) > 0$ ma $A \subset (-\infty, 0]$, e quindi anche in tal caso $\mathbb{P}_2(A) = 0$. Inoltre, per ogni boreliano A ,

$$\begin{aligned} \mathbb{P}_2(A) &= \int_A p_2(x) dx = \int_A p_2(x) \mathbf{1}_{x>0} dx \\ &= \int_A \frac{p_2(x)}{p_1(x)} \mathbf{1}_{x>0} p_1(x) dx = \int_A \frac{p_2(x)}{p_1(x)} \mathbf{1}_{x>0} d\mathbb{P}_1(x), \end{aligned}$$

quindi

$$\frac{d\mathbb{P}_2}{d\mathbb{P}_1}(x) = \frac{p_2(x)}{p_1(x)} \mathbf{1}_{x>0} = \frac{1}{\lambda^2} e^{\lambda x} \mathbf{1}_{0<x<\lambda}.$$

Esercizio 10. a) Si ha

$$\mathbb{Q}(\Omega) = \mathbb{E}(2X) = 2 \cdot \frac{1}{2} = 1$$

quindi $\mathbb{Q}$ è una misura di probabilità.

b) Ovviamente (visto a lezione), $\mathbb{Q} \ll \mathbb{P}$ e $\frac{d\mathbb{Q}}{d\mathbb{P}} = 2X$. Viceversa, preso $A \in \mathcal{F}$ tale che $\mathbb{Q}(A) = 0$ allora $\mathbb{E}(2X\mathbf{1}_A) = 0$. Ma $2X\mathbf{1}_A \geq 0$ $\mathbb{P}$ -q.c., quindi dev'essere $2X\mathbf{1}_A = 0$ $\mathbb{P}$ -q.c. Essendo $\mathbb{P}(X = 0) = 0$, dev'essere necessariamente $\mathbf{1}_A = 0$ $\mathbb{P}$ -q.c., cioè $\mathbb{P}(A) = 0$. Quindi, $\mathbb{P} \ll \mathbb{Q}$ ed allora $\frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{1}{2X}$.

c) Ricordando che $Y \in L^p(\Omega, \mathcal{F}, \mathbb{Q})$ se e solo se $Y \frac{d\mathbb{Q}}{d\mathbb{P}} \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ e in tal caso, $\mathbb{E}^{\mathbb{Q}}(|Y|^p) = \mathbb{E}^{\mathbb{P}}(|Y|^p \frac{d\mathbb{Q}}{d\mathbb{P}})$, si ha:

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}(|e^X|^p) &= \mathbb{E}^{\mathbb{P}}\left(e^{pX} \frac{d\mathbb{Q}}{d\mathbb{P}}\right) = \int_0^{+\infty} e^{px} 2x \cdot 2e^{-2x} dx \\ &= 4 \int_0^{+\infty} x e^{-(2-p)x} dx < \infty \text{ sse } p < 2. \end{aligned}$$

Quindi, $Y \in L^p(\Omega, \mathcal{F}, \mathbb{Q})$ per $p < 2$ e altrimenti, $Y \notin L^p(\Omega, \mathcal{F}, \mathbb{Q})$.

d) Presa Z indipendente da X , si ha, per $\Gamma \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} \mathbb{Q}(Z \in \Gamma) &= \mathbb{E}^{\mathbb{Q}}(\mathbf{1}_{Z \in \Gamma}) = \mathbb{E}^{\mathbb{P}}(\mathbf{1}_{Z \in \Gamma} \cdot 2X) \\ &= \mathbb{E}^{\mathbb{P}}(\mathbf{1}_{Z \in \Gamma}) \cdot \mathbb{E}^{\mathbb{P}}(2X) = \mathbb{E}^{\mathbb{P}}(\mathbf{1}_{Z \in \Gamma}) = \mathbb{P}(Z \in \Gamma) \end{aligned}$$

perché $\mathbf{1}_{Z \in \Gamma}$ e X rimangono indipendenti, da cui la tesi.

Esercizio 11 Dev'essere $p_i \geq 0$ q.o., quindi $a, b \geq 0$, e

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

Ora, $p_1 \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, \text{Leb})$ se e solo se⁴ $\alpha + 1 < 0$ 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 $a = -(\alpha + 1)$. Analogamente, $|x|^p p_1(x) = ax^{\alpha+p} \mathbf{1}_{x>1} \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, \text{Leb})$ se e solo se $\alpha + p < -1$, quindi dev'essere $\alpha < -1 - p$ e in tal caso $\mathbb{E}(|X|^p) = (\alpha + 1)/(\alpha + p + 1)$.

⁴Infatti, indicando con $p_{1n}(x) = p_1(x) \mathbf{1}_{x \in (0, n)}$, allora $0 \leq p_{1n}(x) \uparrow p_1(x)$ per $n \rightarrow \infty$, quindi (da MON) $\int_{\mathbb{R}} p_{1n}(x) dx \uparrow \int_{\mathbb{R}} p_1(x) dx$ per $n \rightarrow \infty$. Ma

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

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

Poi, p_2 è integrabile se e solo se⁵ $\beta > -1$ 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 $|X| \leq 1$ 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 12. a) Poiché $X \in (0, 10)$ q.c., $N = [X] \in E = \{0, 1, \dots, 9\}$. Quindi, $\mathbb{P}(X \leq x, N = n) = 0$ per $x \leq 0$ oppure $n \notin E$. Fissiamo ora $x > 0$ e $n \in E$:

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

Ma

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

e dunque

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

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. In tal caso, fissato $n \in E$,

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

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

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

⁵Posto $p_{2n}(x) = p_2(x) \mathbf{1}_{1/n < |x| < 1}$, allora $0 \leq p_{2n}(x) \uparrow p_2(x)$ q.o. per $n \rightarrow \infty$ e (ancora MON) $\int_{\mathbb{R}} p_{2n}(x) dx \uparrow \int_{\mathbb{R}} p_2(x) dx$. Ora

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

c) Basta prendere $\Gamma = E$: in tal caso $F_n(x) = \int_{-\infty}^x \mathbf{1}_{(n,n+1)}(t)dt$, quindi Q_n ha densità $f_n(x) = \mathbf{1}_{(n,n+1)}(x)$, che corrisponde alla densità di una v.a. $Un(n, n+1)$. 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.

Esercizio 13. a) Lavoriamo con $g = \psi^2$:

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

quindi $g'(\alpha) = 2(\alpha - \mathbb{E}[X]) \geq 0$ se e solo se $\alpha \geq \mathbb{E}(X)$. 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

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

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 $\nabla g = 2(a\mathbb{E}[X^2] - \mathbb{E}[XY] + b\mathbb{E}[X], b - \mathbb{E}[Y] + a\mathbb{E}[X])^t$. L'equazione $\nabla g = 0$ ha come unica soluzione

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

La matrice hessiana

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

è definita positiva⁶, 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 $n \rightarrow +\infty$, (a^*, b^*) è in effetti un punto di minimo assoluto.

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

$$\psi(a^*, b^*) = \left(\mathbb{E}[(Y - a^*X - b^*)^2] \right)^{1/2} = \left(\mathbb{E}[(Y - \mathbb{E}(Y)) - a^*(X - \mathbb{E}(X))]^2 \right)^{1/2}$$

⁶Si ha

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

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

$$= \left(\text{Var}(Y) + a^*{}^2 \text{Var}(X) - 2a^* \text{Cov}(X, Y) \right)^{1/2} = \left(\frac{\text{Var}(X)\text{Var}(Y) - \text{Cov}^2(X, Y)}{\text{Var}(X)} \right)^{1/2}.$$

Inoltre, poiché $\psi^2 \geq 0$, dev'essere $\text{Var}(X)\text{Var}(Y) - \text{Cov}^2(X, Y) \geq 0$, da cui segue la disuguaglianza di Cauchy-Schwarz:

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

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

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”? Il criterio di ottimalità scelto è il seguente: si vogliono determinare due numeri a e b tali da minimizzare è la distanza in L^2 tra la v.a. Y e la v.a. $aX + b$, cioè

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

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

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

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

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

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 $\text{Var}(X)$ 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 legghi 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, com’è ben noto, 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:

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

Ciò significa che la covarianza tra due v.a. è strettamente legata all’unità di misura. Per tale ragione, si introduce la seguente quantità:

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

$\rho_{X,Y}$ 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 ogni⁷ a, b tali che $ab > 0$. Inoltre, abbiamo visto che $\text{Cov}(X, Y)^2 \leq \text{Var}(X)\text{Var}(Y)$, 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.

⁷Se $ab < 0$, 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.