II Università di Roma, Tor Vergata
Dipartimento d'Ingegneria Civile e Ingegneria Informatica
LM in Ingegneria dell'Informazione e dell'Automazione
Complementi di Probabilità e Statistica - Advanced Statistics
Instructors: Roberto Monte & Massimo Regoli
Intermediate Test 2022-12-14

Problem 1 *The scrutiny of group of 100,000 randomly chosen male people in the age 40 – 79 in UK during 2013 – 2015 reveals the following table of average lung cancer incidence*

	smoker	*not smoker*	*total*
lung cancer	10,395	7,407	17,802
not lung cancer	50,078	32,120	82,198
total	60,473	39,527	100,000

Write Ω for the sample space consisting of these 100,000 people and write S [resp. C] for the events of Ω "the people are smokers" [the people are affected by lung cancer]. Let $1_S : \Omega \to \mathbb{R}$ and $1_C : \Omega \to \mathbb{R}$ the indicator functions of the events S and C respectively.

1. *Determine the joint distribution and the joint distribution function of the random vector $(1_S, 1_C)$*
2. *Determine the distributions and the distributions functions of the random variables 1_S and 1_C.*
3. *Are the random variables 1_S and 1_C independent? Compute their correlation.*
4. *What is the probability that a randomly chosen person in Ω is affected by lung cancer, given that he is a smoker [resp. not a smoker]?*
5. *What is the probability that a randomly chosen person in Ω is a smoker [resp. not a smoker], given that he is affected by lung cancer?*
6. *Check the validity of the total probability formula for $\mathbf{P}(S)$ and the Bayes Formula for $\mathbf{P}(C \mid S)$.*

Solution. .

Problem 2 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ a probability space and let $X : \Omega \to \mathbb{R}$ be a real $\mathcal{E}$-random variable uniformly distributed in the interval $[a, b]$. On symbol $X\tilde{\ } Unif(a, b)$. Consider n independent random variables $X_1, \dots, X_n$ with the same distribution of X (a simple random sample of size n drawn from X). Compute the distribution function of the* sample minimum of size n *drawn from X, that is the random variable* $\min(X_1, \dots, X_n) : \Omega \to \mathbb{R}$ *given by*

$$\min(X_1, \dots, X_n)(\omega) \stackrel{def}{=} \min(X_1(\omega), \dots, X_n(\omega)), \quad \forall \omega \in \Omega.$$

Is $\min(X_1, \dots, X_n)$ *an absolutely continuous random variable? Can you compute the density function of* $\min(X_1, \dots, X_n)$*?*

Hint: focus on the representation of the events $\{\min(X_1, \dots, X_n) \leq x\}$ and $\{\min(X_1, \dots, X_n) > x\}$. Hence, focus on the representation of $1 - F_X(x)$ and $(1 - F_X(x))^n$, where $F_X : \mathbb{R} \to \mathbb{R}$ is the distribution function of the random variable X.

Solution. .

Problem 3 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a uniformly distributed random variable with states in the interval $[-1,1]$. In symbols, $X \sim Unif(-1,1)$. Consider the function $g : \mathbb{R} \to \mathbb{R}$ given by*

$$g(x) \overset{def}{=} \begin{cases} 0, & \text{if } x \leq 0. \\ \sqrt{x}, & \text{if } x > 0. \end{cases}$$

1. *Can you show that the function $Y : \Omega \to \mathbb{R}$ given by*

$$Y(\omega) \overset{def}{=} g(X(\omega)), \quad \forall \omega \in \Omega,$$

   is a random variable?
2. *Can you compute the distribution function $F_Y : \mathbb{R} \to \mathbb{R}_+$ of the random variable Y?*
3. *Is Y absolutely continuous?*
4. *Are the first and second order moments of Y finite?*
5. *If the first and second order moments are finite, can you compute $\mathbf{E}[Y]$ and $\mathbf{D}^2[Y]$?*

Solution. .

Problem 4 *Let $F : \mathbb{R}^2 \to \mathbb{R}_+$, briefly F, given by*

$$F(x_1, x_2) \overset{def}{=} \left(1 - e^{-x_1} - e^{-x_2} + e^{-(x_1+x_2)}\right) 1_{\mathbb{R}_+}(x_1)\, 1_{\mathbb{R}_+}(x_2), \quad \forall (x_1, x_2) \in \mathbb{R}^2.$$

Show that F is the distribution function of a real random vector (X_1, X_2) and compute the marginal distribution functions of (X_1, X_2).

1. *Is the function F absolutely continuous?*
2. *Are the entries X_1 and X_2 of the random vector (X_1, X_2) independent random variables?*
3. *Are the entries X_1 and X_2 of the random vector (X_1, X_2) absolutely continuous random variables?*
4. *What is the distribution $F_Z : \mathbb{R}^2 \to \mathbb{R}_+$, briefly F_Z, of the real random variable $Z = \max\{X_1, X_2\}$.*
5. *Is the function F_Z absolutely continuous?*

   Hint: it might be useful to rewrite $F(x_1, x_2)$ in a more convenient form.

Solution. .

Problem 5 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ a probability space, let X and Y be independent standard Bernoulli random variables on Ω. Define $Z \overset{def}{=} X + Y$.*

1. *Compute $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$.*
2. *Are the random variables $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$ uncorrelated?*
3. *Are the random variables $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$ independent?*
4. *By using the properties of the conditional expectation, on account that you are dealing with Bernoulli random variables, can you compute $\mathbf{E}\left[(X+Y)^2 \mid Z\right]$ and $\mathbf{E}[XY \mid Z]$?*

Solution. .

3.1

P3)

$X \sim Unif(-1,1)$ e $g(x) = \begin{cases} 0 & se\ x \le 0 \\ \sqrt{x} & se\ x > 0 \end{cases} = \sqrt{x}\,\mathbb{1}_{(0,+\infty)}$

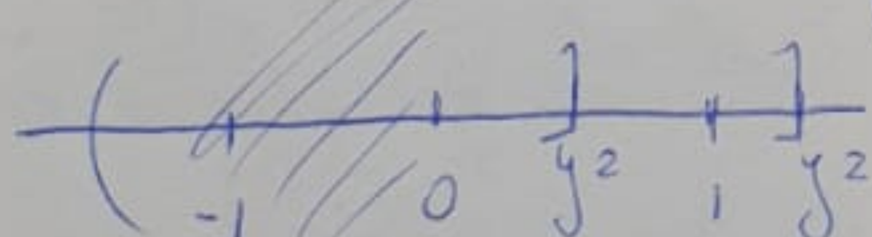
1) Sì $Y = g(X)$ è v.A perché $g(x)$ ~~continua~~ boreliana

2) $\cancel{F_Y(y) = P(Y \le y) = \begin{cases} 0 & se\ X<0 \\ \sqrt{X} & se\ X>0 \end{cases} = \begin{cases} 0 & se \end{cases}}$

$$P(Y \le y) = \begin{cases} 0 & se\ X \le 0 \\ P(\sqrt{X} \le y) & se\ X \ge 0 \end{cases} = \begin{cases} 0 & altrimenti \\ P(X \le y^2) & se\ X > 0 \end{cases}$$

$$\to \int_{(-\infty, y^2]} f_X(x)\, d\mu_L(x) = \int_{(-\infty, y^2]} \frac{1}{2}\, \mathbb{1}_{[0,1]}(x)\, d\mu_L(x) = \frac{1}{2} \int_{(-\infty, y^2] \cap [0,1]} d\mu_L(x) =$$

$$= \frac{1}{2}\, \mu_L\big((-\infty, y^2] \cap [0,1]\big) = \begin{cases} \frac{1}{2} & se\ y \ge 1 \\ \frac{y^2}{2} & se\ 0 < y < 1 \end{cases}$$



$\to 0 < y^2 \le 1 \to y \le \pm 1 \to$ Il caso $y < -1$ e $y > -1$ non ci interessa perché garantito da

$y^2 > 1 \to y > \pm 1$

$$\to F_Y(y) = \frac{1}{2}\, \mathbb{1}_{[1,+\infty)}(y) + \frac{y^2}{2}\, \mathbb{1}_{(0,1)}(y)$$

3.2

3)

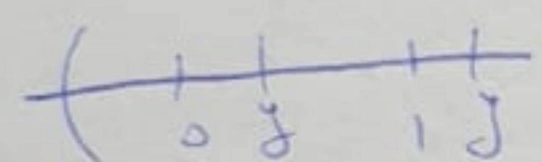
$F'_Y = y\, \mathbb{1}_{(0,1)}(y)$ che è continua ovunque dato che in particolare F_Y è derivabile ovunque

→ Sì è ass. continua

$\cancel{\int_{(-\infty,y]} y\, \mathbb{1}_{(0,1)}(y)\, dy} = \int_{(-\infty,y]} y\, \mathbb{1}_{(0,1)}(y)\, d\mu_L(y) =$

$\underset{\downarrow}{=}$ f continua limitata $\int_{(-\infty,y]\cap(0,1)} y\, dy = \begin{cases} \frac{y^2}{2}\Big|_0^y & \text{se } 0<y<1 \\ \frac{y^2}{2}\Big|_0^1 & \text{se } y\geq 1 \end{cases} = \frac{y^2}{2}\mathbb{1}_{(0,1)} + \frac{1}{2}\mathbb{1}_{(1,+\infty)}$

Sì è ass cont



4)/5)

$E[Y^2] = \int_{\mathbb{R}} y^2 f_Y(y) = \int_{\mathbb{R}} y^3 \mathbb{1}_{(0,1)}(y)\, d\mu_L(y) \underset{\downarrow}{=} \int_0^1 y^3\, dy = \frac{1}{4}$

f cont limitata

→ Ammette momento finito di ordine 2 e quindi anche 1

$E[Y] = \int_{\mathbb{R}} y f_Y(y) = \int_{\mathbb{R}} y^2 \mathbb{1}_{(0,1)}(y)\, d\mu_L(y) = \int_0^1 y^2\, dy = \frac{1}{3}$

→ $D^2[Y] = E[Y^2] - E[Y]^2 = \frac{1}{4} - \frac{1}{9} = \frac{9-4}{36} = \frac{5}{36}$

~~P~~4)

41

$$F_{X_1,X_2} = \left(1 - e^{-x_1} - e^{-x_2} + e^{-(x_1+x_2)}\right) \mathbb{1}_{\mathbb{R}_+}(x_1)\, \mathbb{1}_{\mathbb{R}_+}(x_2)$$

1)

$$\cancel{f_{X_1,X_2}(x_1,x_2)} \rightarrow F'_{X_1,X_2} = \frac{\partial F_{X_1,X_2}}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_2}\left(e^{-x_1} - e^{-(x_1+x_2)}\right) =$$

$$= e^{-(x_1+x_2)} \mathbb{1}_{\mathbb{R}_+}(x_1)\, \mathbb{1}_{\mathbb{R}_+}(x_2)$$

$$\int_{(-\infty,x_1]\times(-\infty,x_2]} F'_{X_1,X_2}\, d\mu_L^2(x_1,x_2) = \int e^{-(u_1+u_2)} \mathbb{1}_{\mathbb{R}_+^2}(u_1,u_2)\, d\mu_L(u_1,u_2) =$$

$$= \int_{\mathbb{R}_+\cap(-\infty,x_1]} du_1 \int_{\mathbb{R}_+\cap(-\infty,x_2]} du_2\, e^{-(u_1+u_2)} = \int_0^{x_1} du_1 \int_0^{x_2} du_2\, e^{-(u_1+u_2)} =$$

f continua limitata + FUBINI + LEBESGUE

$$= \int_0^{x_1} du_1\, e^{-u_1} \int_0^{x_2} e^{-u_2}\, du_2 = -\int_0^{x_1} du_1\, e^{-u_1} \left(e^{-u_2}\right)\Big|_0^{x_2} =$$

$$= -\int_0^{x_1} du_1\, e^{-u_1}\left(e^{-x_2} - 1\right) = \int_0^{x_1} du_1 - e^{-(u_1+x_2)} + e^{-u_1} =$$

$$= \int_0^{x_1} -e^{-(u_1+x_2)}\, du_1 + \int_0^{x_1} e^{-u_1}\, du_1 =$$

4.2

$$= e^{-x_2} \int_0^{x_1} -e^{-u_1} du_1 - \int_0^{x_1} -e^{-u_1} du_1 =$$

$$= e^{-x_2} e^{-u_1}\Big|_0^{x_1} - e^{-u_1}\Big|_0^{x_1} = e^{-x_2}\left(e^{-x_1}-1\right) - \left(e^{-x_1}-1\right) =$$

$$= \left(e^{-(x_2+x_1)} - e^{-x_2} - e^{-x_1} + 1\right) \mathbb{1}_{\mathbb{R}_+}(x_1)\, \mathbb{1}_{\mathbb{R}_+}(x_2) \rightarrow \text{Si è ass cont}$$

2)

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{X_1,X_2}\, d\mu_L(x_2) = \int_{\mathbb{R}} e^{-(x_1+u_2)} \mathbb{1}_{\mathbb{R}_+}(x_1)\, \mathbb{1}_{\mathbb{R}_+}(u_2)\, d\mu_L(u_2) =$$

$$= e^{-x_1} \mathbb{1}_{\mathbb{R}_+}(x_1) \int_0^{+\infty} e^{-u_2} du_2 = e^{-x_1} \mathbb{1}_{\mathbb{R}_+}(x_1) \left(1\right) = e^{-x_1} \mathbb{1}_{\mathbb{R}_+}(x_1)$$

$$f_{X_2}(x_2) = \int_{\mathbb{R}} e^{-(u_1+x_2)} \mathbb{1}_{\mathbb{R}_+}(u_1)\, \mathbb{1}_{\mathbb{R}_+}(x_2)\, d\mu_L(u_1) = e^{-x_2} \mathbb{1}_{\mathbb{R}_+}(x_2) \left(-e^{-u_1}\right)_0^{+\infty} =$$

$$= e^{-x_2} \mathbb{1}_{\mathbb{R}_+}(x_2)$$

$$\rightarrow f_{X_1} f_{X_2} = e^{-(x_1+x_2)} \mathbb{1}_{\mathbb{R}_+}(x_1)\, \mathbb{1}_{\mathbb{R}_+}(x_2) = f_{X_1,X_2} \rightarrow \text{Si sono indipendenti}$$

4.3

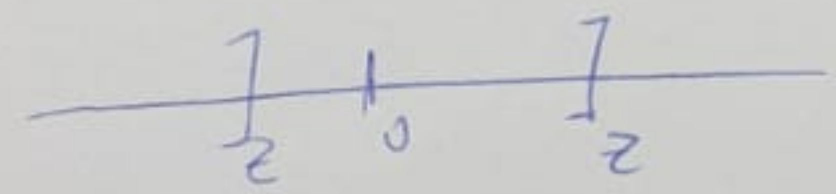
3) Si perché f_{X_1} e f_{X_2} continue e limitate

4) $Z = \max\{X_1, X_2\}$

$$P(Z \leq z) = P(X_1 \leq z, X_2 \leq z) \overset{X_1, X_2 \text{ ind}}{=} P(X_1 \leq z)\, P(X_2 \leq z)$$

$$\{Z \leq z\} = \{X_1 \leq z, X_2 \leq z\}$$

$$P(X_1 \leq z) = \int_{(-\infty, z]} f_{X_1}(x_1)\, d\mu_L(x_1) = \int_{(-\infty, z)} e^{-x_1} \mathbb{1}_{\mathbb{R}_+}(x_1)\, d\mu_L(x_1) =$$



$$= \int_0^z e^{-x_1}\, dx_1 = -e^{-x_1}\Big|_0^z = 1 - e^{-z} \mathbb{1}_{\mathbb{R}_+}(z)$$

$$P(X_2 \leq z) = P(X_1 \leq z)$$

$$\rightarrow P(Z \leq z) = 1 + e^{-2z} - 2e^{-z} \mathbb{1}_{\mathbb{R}_+}(z)$$

$F'_Z = -2e^{-2z} + 2e^{-z} \mathbb{1}_{\mathbb{R}_+}(z)$ che è derivabile ovunque quindi si

4.4

$\cancel{\int_{(-\infty, x_1]} e^{-x_1} 1_{\mathbb{R}_+} d\mu_1(x_1) = \int_0^{x_1} e^{-x_1} =}$

5)

$$\int_{(0,z]} 2e^{-z} - 2e^{-2z} \, dz = -2e^{-z}\Big|_0^z + e^{-2z}\Big|_0^z =$$

$$= -2\left(e^{-z} - 1\right) + \left(e^{-2z} - 1\right) =$$

$$\cancel{= \left(1 \cancel{2} - 2e^{-z} - \cancel{1}\right) + e^{-2z} 1_{\mathbb{R}_+}(z) =}$$

$$= -2e^{-z} + \cancel{2}^{1} + e^{-2z} \cancel{1} \quad - \cancel{\text{ssss}} \ S1 \longrightarrow$$

$\cancel{\int 2e^{-2z} + \int 2e^{-z}}$

$\cancel{= e^{-2z} \qquad = -2e^{-z}}$

P9)

5.1

$X, Y \sim Ber(p)$ indipendenti

$$X = \begin{cases} 0 & \text{con } q = 1-p \\ 1 & \text{con } p \end{cases}$$

$$Z = X+Y = \begin{cases} 2 & \text{se } X=1, Y=1 \text{ con } p^2 \\ 1 & \text{se } X=0, Y=1 \text{ OR } X=1, Y=0 \text{ con } 2pq \\ 0 & \text{se } X=0, Y=0 \text{ con } q^2 \end{cases}$$ cioè $Z \sim Bin(2, p)$ dato X e Y indipendenti

1) $E[X|Z] = \sum_{k \in \{0,1,2\}} E[X|Z=k] = E[X|Z=0] + E[X|Z=1] + E[X|Z=2]$

$$E[X|Z=0] = \frac{1}{P(Z=0)} \int_{\{Z=0\}=\{X=0, Y=0\}} X \, dP = \emptyset$$

$$E[X|Z=1] = \frac{1}{P(Z=1)} \int_{\{Z=1\}=\{X=0,Y=1\} \cup \{X=1,Y=0\}} X \, dP = \frac{1}{2pq} 0 \cdot pq + 1 \cdot pq = = \frac{1}{2}$$

$$E[X|Z=2] = \frac{1}{P(Z=2)} \int_{\{Z=2\}=\{X=1,Y=1\}} X \, dP = \frac{1}{p^2}(p^2) = 1$$

$$\rightarrow E[X|Z] = \frac{1}{2} \mathbb{1}_{\{Z=1\}} + \mathbb{1}_{\{Z=2\}} = \frac{1}{2} Z$$

anche $E[Y|Z] = \frac{1}{2} Z$

5.2

2) $Cov\left(E[X|Z], E[Y,Z]\right) =$

$$= \underbrace{E\left[E[X|Z]\,E[Y|Z]\right]}_{E_1} - \underbrace{E\left[E[X|Z]\right]\cdot E\left[E[Y|Z]\right]}_{E_2} = 0 \,?$$

$$E_1 = E\left[\frac{1}{2}Z\,\frac{1}{2}Z\right] = E\left[\frac{1}{4}Z^2\right] = \frac{1}{4}E[Z^2] = \frac{1}{4}E[(X+Y)^2] =$$

$$= \frac{1}{4}\left(E[X^2] + E[Y^2] + 2E[XY]\right) = \frac{1}{4}\left(E[X] + E[Y] + 2E[X]\,E[Y]\right) =$$

$X, Y \sim Ber(p)$
e indipendenti

$$= \frac{p}{4} + \frac{p}{4} + \frac{p^2}{2} = \frac{p}{2} + \frac{p^2}{2}$$

$$E_2 = \frac{1}{\cancel{2}}\cancel{2}p\,\frac{1}{\cancel{2}}\cancel{2}p = p^2$$

$\rightarrow E_1 - E_2 = \frac{p}{2} + \frac{p^2}{2} - p^2 = \frac{p}{2} - \frac{p^2}{2} = \frac{p}{2}(1-p) \neq 0$ sono correlate

$\hookrightarrow E[X|Z] = \frac{1}{2}Z$ con $Z \sim Bin(2, p)$

3) ↓
NON SONO
INDIPENDENTI

~~E[(X+~~ 4)

$$E[(X+Y)^2|Z] = E[X^2+Y^2+2XY|Z] =$$

$X,Y \sim Ber(p) \to XY \sim Ber\, p$ e $Y^2, X^2 \sim Ber(p)$

$$\underset{\text{LINEARITÀ}}{=} E[X^2|Z] + E[Y^2|Z] + 2E[XY|Z] =$$

$$= E[X|Z] + E[Y|Z] + 2E[X|Z] = 4\, E[X|Z] = 4 \cdot \frac{1}{2} Z = 2Z$$

con $Z \sim Bin(2, p)$

e $E[XY|Z] = E[X^2|Z] = E[X|Z] \sim \frac{1}{2} Z$ con $Z \sim Bin(2, p)$