(30L)

7/07/2014

# Secondo compitino di
# **Probabilità e Statistica**
## Laurea Triennale in Matematica
## 17/06/2014

**COGNOME e NOME** ....RIGON DAVIDE....

**N. MATRICOLA**....1070080....

**Esercizio 1.** (V. 3 punti.)
Date due variabili $X$ e $Y$ in $L^2$. Dimostrare che se $X$ e $Y$ sono indipendenti allora sono non correlate.

Se $X, Y$ sono indipendenti e sono in $L^2$ allora $XY \in L^1$ perché $|XY| \leq \frac{1}{2}(X^2 + Y^2)$ e per la linearità del valore medio si conclude che $XY \in L^1$.
Inoltre, sono indipendenti $X$ e $Y$ se e solo se (per definizione) per ogni $A, B \subseteq \Omega$ si ha $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(X \in B)$. ove $\Omega$ è lo spazio campionario.
Quindi si ha

$$\mathbb{E}[XY] = \sum_{x \in \Omega} \sum_{y \in \Omega} xy\, \mathbb{P}(X = x, Y = y)$$

$$= \sum_{x \in \Omega} \sum_{y \in \Omega} x\, \mathbb{P}(X = x) \cdot y\, \mathbb{P}(Y = y)$$

$$= \sum_{x \in \Omega} x\, \mathbb{P}(X = x) \sum_{y \in \Omega} y\, \mathbb{P}(Y = y)$$

$$= \left(\sum_{x \in \Omega} x\, \mathbb{P}(X = x)\right) \left(\sum_{y \in \Omega} y\, \mathbb{P}(Y = y)\right)$$

$$= \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

e quindi se ne deduce che, visto che $Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, $Cov(X,Y) = 0$ cioè $X, Y$ sono non correlate.

**Esercizio 2.** (V. 3 punti.)
Esporre la proprietà 'assenza di memoria' per variabili aleatorie esponenziali.

le variabili aleatorie sono assolutamente continue, con ~~densità~~ densità $\lambda e^{-\lambda x}$, $\forall x \in [0, +\infty)$.
la proprietà di assenza di memoria per esponenziali, analoga a quella per le variabili discreta di tipo geometrica, afferma che

$\forall s, t \in \mathbb{R}$ $s, t \geq 0$, si ha

$\mathbb{P}(X > s+t \mid X > s) = \mathbb{P}(X > t)$ se $X$ è v.a. assolutamente continua ed esponenziale.

Si può dimostrare ricordando che $\mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t)$ e ricordando $\mathbb{P}(X \leq t) = 1 - e^{-\lambda t}$ $\forall t \geq 0$, se $X$ è esponenziale, ricordando inoltre come funziona la probabilità condizionata, e le proprietà dell'esponenziale ($e^{x+y} = e^x \cdot e^y$)

**Esercizio 3.** (V. 10 punti.)
Sia $(X_n)_{n\in\mathbb{N}}$ una successione di variabili aleatorie indipendenti ed identicamente distribuite sullo spazio di probabilità $(\Omega, \mathcal{A}, P)$. Supponiamo che le v.a. $X_n$ abbiano distribuzione uniforme sull'intervallo $(1,3)$, $X_n \sim Unif(1,3)$. Per ogni $n \in \mathbb{N}$ siano definite le seguenti variabili aleatorie:

$$Y_n := \sum_{i=1}^{n} \frac{X_i}{2^{n-i}} \qquad W_n := \sum_{i=1}^{n} \frac{Y_i}{n}$$

Calcolare media e varianza di $Y_n$. Le variabili aleatorie $(Y_n)_{n\in\mathbb{N}}$ sono indipendenti? Sono non correlate? Dimostrare che vale la seguente uguaglianza:

$$P(\{\omega \in \Omega \mid \lim_{n\to\infty} W_n(\omega) = 4\}) = 1$$

(`Può essere utile l'uguaglianza` $\sum_{j=n}^{m} \alpha^j = \frac{\alpha^n - \alpha^{m+1}}{1-\alpha}$).

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\sum_{i=1}^{n} \frac{X_i}{2^{n-i}}\right] = \mathbb{E}\left[\sum_{i=1}^{n} \frac{2^i X_i}{2^n}\right] = \frac{1}{2^n}\mathbb{E}\left[\sum_{i=1}^{n} 2^i X_i\right]$$

$$= \frac{1}{2^n} \cdot \sum_{i=1}^{n} 2^i \mathbb{E}[X_i] = \frac{1}{2^n}\cdot 2\sum_{i=1}^{n} 2^{i} = \frac{2}{2^n}\cdot\frac{2-2^{n+1}}{1-2}$$

$$= 4\left(\frac{2^n-1}{2^n}\right)$$

$$Var(Y_n) = Var\left(\sum_{i=1}^{n}\frac{2^i X_i}{2^n}\right) = \frac{1}{4^n} Var\left(\sum_{i=1}^{n} 2^i X_i\right)$$

$$= 1/4^n \cdot \left(\sum_{i=1}^{n} 4^i Var(X_i)\right)$$

$$= \frac{1}{3\cdot 4^n}\cdot\sum_{i=1}^{n} 4^i = \frac{4}{9}\cdot\frac{4^n-1}{4^n}$$

considerando che $\mathbb{E}[X_i] = \frac{3+1}{2} = 2$ $\forall i$ e $Var(X_i) = \frac{(3-1)^2}{12} = 1/3$ $\forall i$ visto che sono uniformi su $(1,3)$. Inoltre ho utilizzato le note proprietà di varianza e valore medio.

le variabili aleatorie $(Y_n)_{n\in\mathbb{N}}$ ~~non~~ sono correlate: infatti

$$Cov(Y_1, Y_2) = Cov\left(X_1, \frac{X_1}{2} + X_2\right) = \mathbb{E}\left[X_1\cdot\left(\frac{X_1}{2}+X_2\right)\right] - \mathbb{E}[X_1]\cdot\mathbb{E}\left[\frac{X_1}{2}+X_2\right]$$

$$= \frac{1}{2}\mathbb{E}[X_1^2] + \mathbb{E}[X_1]\,\mathbb{E}[X_2] - \frac{1}{2}(\mathbb{E}[X_1])^2 - \mathbb{E}[X_1]\,\mathbb{E}[X_2]$$



$$= \frac{1}{2}\left(\mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2\right) = \frac{1}{2} Var(X_1) = \frac{1}{6} \neq 0$$

e quindi come conseguenza non sono nemmeno indipendenti.

In realtà si potrebbe dimostrare per due variabili generali $Y_n, Y_{n+2}$ che sono correlate ma è più complicato ed inoltre bastava solo fornire un esempio concreto per verificare che son correlate le $(Y_n)_{n\in\mathbb{N}}$

Quello che si chiede di dimostrare è la legge forte dei grandi numeri, che in questo caso non può essere applicata perché le $(W_n)_{n\in\mathbb{N}}$ non sono i.i.d. in quanto $\mathbb{E}[W_n]$ dipenderà ovviamente da $n$.

Bisogna procedere per un'altra strada.

$$W_n = \sum_{i=1}^{n} \frac{Y_i}{n} = \frac{Y_1 + \dots + Y_n}{n} = \frac{X_1 + (\frac{X_1}{2} + X_2) + (\frac{X_1}{4} + \frac{X_2}{2} + X_3) + \dots + (\frac{X_1}{2^{n-1}} + \frac{X_2}{2^{n-2}} + \dots + X_n)}{n}$$

cioè $$W_n = \frac{X_1\left(1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{n-1}}\right) + X_2\left(1 + \frac{1}{2} + \dots + \frac{1}{2^{n-2}}\right) + \dots + X_{n-1}\left(1 + \frac{1}{2}\right) + X_n(1)}{n}$$

devo verificare che $\lim_{n\to\infty} W_n = 4$ quasi certamente.

Considero $X_1$: $$1 + \frac{1}{2} + \dots + \frac{1}{2^{n-1}} = \sum_{i=0}^{n-1} \frac{1}{2^i} = \frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}} = \left(\frac{2^n - 1}{2^n}\right) \cdot 2 = 2 - \frac{1}{2^{n-1}}$$

$X_2$: $$1 + \frac{1}{2} + \dots + \frac{1}{2^{n-2}} = \frac{1 - \frac{1}{2}^{n-1}}{1 - \frac{1}{2}} = \left(\frac{2^{n-1} - 1}{2^{n-1}}\right) \cdot 2 = 2 - \frac{1}{2^{n-2}}$$

$\vdots$

$X_{n-1}$: $$1 + \frac{1}{2} = \left(\frac{2^2 - 1}{2^2}\right) \cdot 2 = 2 - \frac{1}{2^1}$$

$X_n$: $$1 = \left(\frac{2^1 - 1}{2}\right) \cdot 2 = 2 - \frac{1}{2^0}$$

e posso quindi raccogliere e spezzare il limite, ottenendo

$$\lim_{n\to\infty} W_n = \lim_{n\to\infty} \frac{2(X_1 + \dots + X_n)}{n} - \lim_{n\to\infty} \frac{\frac{X_1}{2^{n-1}} + \frac{X_2}{2^{n-2}} + \dots + \frac{X_n}{2^0}}{n}$$

Ora il secondo ~~membro~~ termine banalmente tende a zero quando $n \to +\infty$, mentre per quanto concerne il primo termine, si si deve ricordare che la successione $(X_n)_{n\in\mathbb{N}}$ è i.i.d. con stessa media e varianza e che quindi soddisfa la legge forte dei grandi numeri, ossia $\lim_{n\to\infty} \frac{X_1 + \dots + X_n}{n} = \mathbb{E}[X_i] = 2$ quasi certamente

e quindi per la linearità del limite



$$\lim_{n\to\infty} W_n = 4 - 0 = 4$$ quasi certamente, cioè

$$\mathbb{P}\left(\{w \in \Omega : \lim_{n\to\infty} W_n(w) = 4\}\right) = 1.$$

**Esercizio 4.** (V. 8 punti.)
Un'azienda produce barrette di cioccolato. Consideriamo un insieme costitutito da $n = 25$ barrette. Indichiamo con $(X_n)_{n\in\{1,\dots,25\}}$ il loro peso (in grammi) e supponiamo che si distribuisca come una distribuzione normale di media $\mu$ e varianza $\sigma^2$, con $\mu$ e $\sigma$ non noti. (Assumiamo che le v.a. siano indipendenti). Indichiamo con $\overline{X} := \sum_{i=1}^{25} \frac{X_i}{n}$ la media campionaria, con $S^2 := \sum_{i=1}^{25} \frac{(X_i-\overline{X})^2}{n-1}$ la varianza campionaria, e sia infine $T := \sum_{i=1}^{25} X_i^2$. Supponiamo che si realizzi $X_i(\omega) = x_i$, per $i \in \{1,\dots,25\}$ con $\sum_{i=1}^{25} x_i = 2475$ e $\sum_{i=1}^{25} x_i^2 = 245625$.

(a) Determinare un intervallo di confidenza centrato per $\mu$ con livello di confidenza $\gamma = 95\%$.
(b) Determinare un intervallo di confidenza destro per $\mu$ con livello di confidenza $\gamma = 99\%$.

Prima di determinare gli intervalli, trovo i valori di $\overline{X}$ e $S^2$.

$$\overline{X} = \sum_{i=1}^{25} \frac{X_i}{n} = \frac{\sum_{i=1}^{25} X_i}{n} = \frac{2475}{25} = 99$$

$$S^2 = \sum_{i=1}^{n} \frac{(X_i-\overline{X})^2}{n-1} = \frac{1}{24}\left(\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i\overline{X} + \sum_{i=1}^{n} \overline{X}^2\right)$$

$$= \frac{1}{24}\cdot\left(245625 - 2\cdot 99\cdot 2475 + 25\cdot 99^2\right)$$

$$= 25 \quad \text{cioè} \quad S = 5$$

(a) $\cancel{\alpha = 1-\gamma = 5\% = 0,05}$

utilizzo il fatto che $P(T > t_\alpha) = \alpha$ se $T$ è t-di Student

Ho/cioè che $P\left(\left|\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\right| < t_\alpha\right) = 1-2\alpha = 95\% = 0,95$

~~[illegible]~~ $\cancel{\beta = \frac{\alpha}{2} = 0,025}$ ~~ho~~

$$\cancel{P\left(\left|\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\right| < t_\beta\right) = 1-\beta =}$$

da cui ricavo $\alpha = \frac{1-0,95}{2} = 0,025$

cioè $t_\alpha = 2,064$ perché confrontato sulle Tavole con valore $n-1 = 24$

Cioé l'intervallo da considerare é

$$\bar{X} - \frac{s}{\sqrt{n}} t_\alpha \leq \mu \leq \bar{X} + \frac{s}{\sqrt{n}} t_\alpha \quad \text{cioé}$$

$$\mu \in [\, 96,936 \,;\, 101,064 \,]$$

(b) Per gli stessi motivi ma considerando ora solo l'intervallo destro si ha

$$\mathbb{P}\left( \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \leq t_\alpha \right) = 1 - \alpha \quad \forall \alpha \in (0,1)$$

$$= 99\%$$

$$= 0,99$$

da cui si ricava $\alpha = 0,01$ e quindi l'intervallo é

$$\mu \geq \bar{X} - t_{0,01} \cdot \frac{s}{\sqrt{n}} = 99 - 2,492 \cdot 1 = 96,508$$

cioé

$$\mu \in [\, 96,508, \, +\infty)$$

in quanto $s = 5 = \sqrt{n}$ cioé $\frac{s}{\sqrt{n}} = 1$

**Esercizio 5.** (V. 6 punti.)
Sia $X$ e $Y$ due variabili aleatorie. Supponiamo che $Y$ abbia distribuzione uniforme sull'intervallo $(0,1)$, mentre $X$ sia una variabile aleatoria assolutamente continua con densità $f_X$ data da:

$$f_X(x) = \begin{cases} \frac{1}{x} & x \in (1, \alpha) \\ 0 & \text{altrimenti} \end{cases}$$

per una qualche costante $\alpha > 1$
(a) Determinare $\alpha$.
(b) Calcolare la funzione di ripartizione di $X$.
(c) Trovare una funzione $g : (0,1) \to \mathbb{R}$ crescente tale che posto $T := g(Y)$ si abbia $X \sim T$.

(a) essendo $f_X(x)$ una densitá di una variabile aleatoria ~~di~~ assolutamente continua, deve valere $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$. Ora questo integrale puó essere considerato solo tra $1$ e $\alpha$, perché nei restanti casi $f_X(x) = 0$ e quindi si ha $1 = \int_1^\alpha \frac{1}{x}\,dx = \ln x \Big|_{x=1}^{x=\alpha} = \ln\alpha - \ln 1 = \ln\alpha$ cioè

$\alpha = e$

(b) Per la funzione di ripartizione, per definizione, basta integrare la densitá, ricordandosi che

$F_X(t) = 0$ se $t < 1$

$F_X(t) = \int_1^t \frac{1}{x}\,dx = \ln t$ se $1 \le t < e$

$F_X(t) = \int_1^e \frac{1}{x}\,dx + \int_e^t 0\,dx = 1$ se $t \geq e$ cioe

$$F_X(t) = \begin{cases} 0 & t < 1 \\ \ln t & 1 \le t < e \\ 1 & t \geq e \end{cases}$$

e infatti $F_X$ é continua

(c) Per un teorema ~~dim~~ visto in classe la funzione in questione è proprio la pseudo-inversa di $X$ cioè visto che $F_X = \begin{cases} 0 & \text{se } t < 1 \\ \ln t & \text{se } 1 \leq t < e \\ 1 & \text{se } t \geq e \end{cases}$

allora la pseudo-inversa coincide con l'inversa locale in $[1, e)$ cioè si avrà

$$g: (0,1) \longrightarrow \mathbb{R} \qquad y \longrightarrow e^y$$

che tra l'altro è crescente.

Posto infatti $T = g(Y)$ con $Y \sim Unif(0,1)$ si ha

$$F_T(t) = \mathbb{P}(T \leq t) = \mathbb{P}(e^Y \leq t) = \begin{cases} 0 & \text{se } t < 1 \\ \ln t & \text{se } 1 \leq t < e \\ 1 & \text{se } t \geq e \end{cases}$$

perché $Y$ prende valori solo nell'intervallo $(0,1)$.