MS 25/09/25

ESERCIZI SU DENSITÀ E DAGS

Per noi i dati sono realizzazioni $x_1, \dots x_n$ (valori osservati, istanze) di variabili aleatorie $X_1, \dots X_n$.

La densità congiunta di un vettore aleatorio $\underline{X} = (X_1, \dots X_n)$ è

$$f(x_1, \dots x_n)$$ (densità) congiunta

Nel caso discreto, $f(x_1, \dots x_n)$ descrive la distribuzione di masse di probabilità

$$P(\underline{X} \in A) = \sum_{x \in A} f(x_1, \dots x_n)$$

Nel caso continuo invece

$$P(\underline{X} \in A) = \int_A f(x_1, \dots x_n)\, dx_1 \dots dx_n$$

Per esempio $n=1$, $X \sim \text{Binomiale}(10, \frac{1}{3})$

$$P(X \leq 4) = \sum_{i=1}^{4} \binom{10}{i} \left(\frac{1}{3}\right)^i \left(1 - \frac{1}{3}\right)^{10-i}$$

Quando $n=2$, scriveremo (X, Y) invece di (X_1, X_2) e definiamo

(densità) marginale

$$f_X(x) = \sum_y f(x, y) \quad \text{nel caso discreto}$$

$$f_X(x) = \int_y f(x, y)\, dy \quad \text{" " continuo}$$

Con ovvie generalizzazioni a $n > 2$:
per esempio,

$$f_{XY}(x,y) = \int_z f(x,y,z)\, dz$$

$$f_X(x) = \int_y \int_z f(x,y,z)\, dy\, dz$$

e così via.

Le densità condizionate (o subordinate) sono

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

(densità) marginale di Y dato $X=x$ (condizionatamente a $X=x$)

e così via.

Teorema di Bayes (per v.a. continue)

$$f_{X|Y}(x,y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x)\, f_{Y|X}(y|x)}{\int f(x',y)\, dx'}$$

$$= \frac{f_X(x)\, f_{Y|X}(y|x)}{\int f_X(x')\, f_{Y|X}(y|x')\, dx'}$$

definizione di densità condizionata

e così via.

Tornando all'esercizio di ieri:

1. Consider a binary screening situation formulated in terms of the following random variables:

   - D = indicator whether a randomly sampled person is diseased
   - T = indicator whether a single test given to the person is positive
   - T_1, T_2 = similar indicators for two conditionally independent tests

   and compute the following probabilities.

   (a) Consider a single test. Assuming:
   - $\mathrm{P}(\mathrm{D}=1) = 0.01$ (prevalence),
   - $\mathrm{P}(\mathrm{T}=1|\mathrm{D}=1) = 0.90$ (sensitivity)
   - and $\mathrm{P}(\mathrm{T}=0|\mathrm{D}=0) = 0.95$ (specificity),

   compute
   $$\mathrm{P}(\mathrm{D}=1|\mathrm{T}=1) \quad \text{(positive predictive value)}$$

   (b) Consider repeating two conditionally independent tests on the same person. Assume sensitivity and specificity of T_1 and T_2 are the same as T. Compute
   $$\mathrm{P}(\mathrm{D}=1|\mathrm{T}_1=1, \mathrm{T}_2=1).$$

$$f_D(1) = 0.01$$

$$f_{T|D}(1|1) = 0.90$$

$$f_{T|D}(0|0) = 0.95$$

perciò

$$f_{D|T}(1|1) = \frac{f_{D,T}(1,1)}{f_T(1)} = \frac{f_D(1)\, f_{T|D}(1|1)}{\sum_{i=0}^{1} f_D(i)\, f_{T|D}(1|i)} = \frac{0.01 \times 0.90}{0.99(1-0.95) + 0.01 \times 0.90} = 0.154$$

2)

Se T_1 e T_2 sono condizionatamente indipendenti, allora

$$P(T_1=i \cap T_2=j \mid D=1) = P(T_1=i \mid D=1)\,P(T_2=j \mid D=1)$$

e

$$P(T_1=i \cap T_2=j \mid D=0) = P(T_1=i \mid D=0)\,P(T_2=j \mid D=0)$$

e potete calcolare

def di densità condizionata

$$f_{D|T_1,T_2}(1|1,1) = \frac{f_D(1)\, f_{T_1T_2|D}(1,1|1)}{f_{T_1T_2}(1,1)}$$

indipendenza condizionata

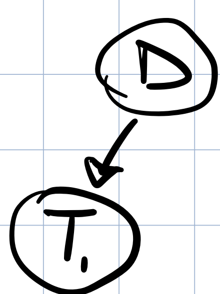
$$= \frac{f_D(1)\, f_{T_1|D}(1|1)\, f_{T_2|D}(1|1)}{f_{T_1T_2}(1,1)}$$

$$= \frac{0.01 \times 0.90 \times 0.90}{0.99 \times 0.05 \times 0.05 + 0.01 \times 0.90 \times 0.90} = 0.766$$

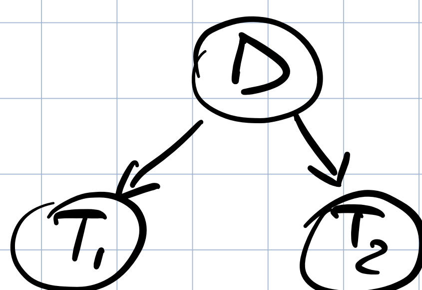
$\Downarrow$

0.154

Viene naturale rappresentare 1) come



e la situazione 2)



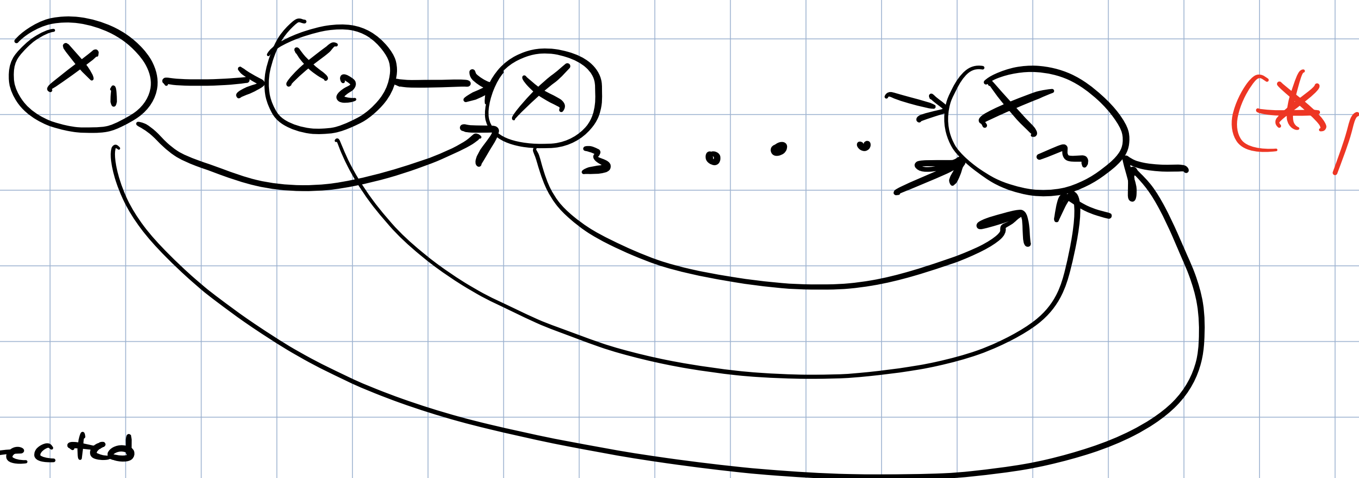
Formalizziamo questa idea.
Dato un vettore $(X_1, \dots X_n)$, possiamo scriverne la congiunta

$$f(x_1 \dots x_n) = f_{X_1 \dots X_{n-1}}(x_1 \dots x_{n-1}) f_{X_n | X_1 \dots X_{n-1}}(x_n | x_1 \dots x_{n-1})$$

$$= f_{X_1 \dots X_{n-2}}(x_1 \dots x_{n-2}) * f_{X_{n-1} | X_1 \dots X_{n-2}}(x_{n-1} | x_1 \dots x_{n-2}) *$$
$$f_{X_n | X_1 \dots X_{n-2} X_{n-1}}(x_n | x_1 \dots x_{n-1} x_{n-2})$$

...

$$= f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_1 X_2}(x_3 | x_1 x_2) \dots$$
$$\dots f_{X_n | X_1 \dots X_{n-1}}(x_n | x_1 \dots x_{n-1})$$



Directed
Acyclic
Graph

Ora, può succedere che alcuni archi siano mancanti, semplificando la descrizione della congiunta

ESEMPIO 1

$$X_1 \to X_2 \to X_3 \; \dots \; X_{n-1} \to X_n \qquad (*)$$

$$f(x_1 \dots x_n) = \prod_{i=1}^{n} f_{X_i|X_{i-1}}(x_i|x_{i-1})$$

dove $f_{X_1|\underset{\emptyset}{X_0}}(x_1|\underset{\emptyset}{\cancel{x_0}}) = f_{X_1}(x_1)$

Questa è chiamata catena di Markov.

ESEMPIO 2: v.a. indipendenti

$$X_1 \quad X_2 \quad \dots \quad X_n \qquad (*)$$

$$f(x_1 \dots x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

ESEMPIO 3: v.a. i. i. d. come $X \sim f_X(x)$

$$X_1 \quad X_2 \quad \dots \quad X_n \qquad (*)$$

$$f(x_1 \dots x_n) = \prod_{i=1}^{n} f_X(x_i)$$

ESEMPIO 4 : dall'esercizio 1

$$\begin{tikzcd} & D \arrow[dl] \arrow[dr] & \\ T_1 & & T_2 \end{tikzcd}$$

(a volte chiamata connessione a forchetta, the FORK)

$$f(d, t_1, t_2) = f_D(d)\, f_{T_1|D}(t_1|d)\, f_{T_2|D}(t_2|d)$$

*

In generale, una Rete Bayesiana (Bayesian Network, BN) è

BN = DAG + fattorizzazione della congiunta

↳ come in *

E' ovvio che sia una "Rete" (i nodi rappresentano componenti di un vettore aleatorio).

"Bayesiano" deriva dal fatto che per calcolare probabilità condizionate di nodi "profondi" * dati nodi a margine si usa Bayes.

Ma la cosa più importante è la fattorizzazione, cioè il concetto base l'indipendenza condizionata.

* Nell' esempio 4,
D è profondo
e T_1 e T_2 sono foglie,
margini, dell' albero.
Per calcolare $f_{D|T_1 T_2}(\cdot|\cdot\cdot)$
usiamo Bayes.

ESEMPIO

$$X_1 \longrightarrow X_2 \longrightarrow X_3$$

(catena di markov)

Sussiste l'indipendenza condizionata

$$\underbrace{X_1 \perp\!\!\!\perp X_3}_{\text{"X_1 e X_3 sono condizionatamente indipendenti dato X_2"}} \mid X_2$$

Perché, siccome vige la fattorizzazione

$$f(x_1, x_2, x_3) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_2}(x_3|x_2)$$

abbiamo

$$f_{X_1,X_3|X_2}(x_1, x_3|x_2) = \frac{f(x_1, x_2, x_3)}{f_{X_2}(x_2)}$$

$$= \frac{f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_2}(x_3|x_2)}{f_{X_2}(x_2)}$$

$$= \frac{f_{X_1|X_2}(x_1|x_2) \cancel{f_{X_2}(x_2)} f_{X_3|X_2}(x_3|x_2)}{\cancel{f_{X_2}(x_2)}}$$

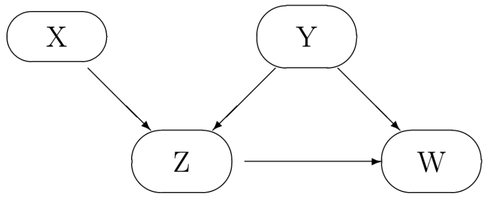
$$= f_{X_1|X_2}(x_1|x_2) f_{X_3|X_2}(x_3|x_2)$$

Problema interessante: quali sono le relazioni di indipendenza condizionata indotte da un BN?

$\longrightarrow$ d-separation

ESECIZIO 3

3. Let X, Y, Z and W be binary variables with a joint distribution represented by the following DAG



and such that

- $P(X = 1) = 0.7$; $P(Y = 1) = 0.3$;
- $P(Z = 1|X = 0 \cap Y = 0) = 0.2$; $P(Z = 1|X = 0 \cap Y = 1) = 0.3$;
- $P(Z = 1|X = 1 \cap Y = 0) = 0.2$; $P(Z = 1|X = 1 \cap Y = 1) = 0.4$;
- $P(W = 1|Y = 0 \cap Z = 0) = 0.2$; $P(W = 1|Y = 0 \cap Z = 1) = 0.3$;
- $P(W = 1|Y = 1 \cap Z = 0) = 0.2$; $P(W = 1|Y = 1 \cap Z = 1) = 0.4$.

(a) Compute the probability that at least one variable equals 1.

(b) Having observed $Y = 0$ and $Z = 0$, compute the probability that $X = 1$.

La cosa più semplice è calcolare la congiunta tramite fattorizzazione

x	y	z	w
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

$f(x, y, z, w)$

$0.3 \times 0.7 \times 0.8 \times 0.8$

$0.3 \times 0.7 \times 0.8 \times 0.2$

...

Una volta calcolata la congiunta, si possono fare tutti i calcoli di probabilità.