JNOTES

CREATIVE NOTES

PROBABILITÀ

CONCETTO CHIAVE

LA PROBABILITÀ MODELLA FENOMENI COMPLESSI IN CUI L'ESITO È INCERTO A CAUSA DI TROPPI FATTORI INFLUENZANTI
(ES SE LANCIO UNA MONETA SEMPRE NELLE STESSE CONDIZIONI, STESSA FORZA, TUTTO IL RISULTATO SARÀ CERTO USCIRÀ SEMPRE LO STESSO VALORE. IN QUESTO CASO NO)

EVENTO DETERMINISTICO: ESITO PREVEDIBILE CON CERTEZZA.
ESEMPIO: LASCIO CADERE UNA PALLA DA UN ALTEZZA DI 10 m SULLA TERRA, TOCCHERÀ IL SUOLO (FISICA CLASSICA).

EVENTO ALEATORIO: ESITO INCERTO.

ESEMPIO: "DOMANI PIOVERÀ" DIPENDE DA FATTORI NON CONTROLLABILI

PERCHÉ USARE LA PROBABILITÀ?
PRENDERE DECISIONI IN CONDIZIONI DI INCERTEZZA (ES PREVISIONI METEO o DIAGNOSI MEDICHE)

ESISTONO 3 DIVERSI TIPI DI INTERPRETAZIONI DELLA PROBABILITÀ

a) INTERPRETAZIONE CLASSICA

$$\text{PROBABILITÀ} = \frac{\text{CASI FAVOREVOLI}}{\text{CASI POSSIBILI}}$$

(ASSUME EVENTI EQUIPROBABILI)
NON INFLUENZATI DAL PASSATO CON STESSA PROBABILITÀ DI ACCADERE

ES: LANCIO DI UN DADO EQUO (NON TRUCCATO)
$\rightarrow P(3) = \frac{1}{6}$ = PROBABILITÀ CHE ESCA 3

b) INTERPRETAZIONE FREQUENTISTA

PROBABILITÀ = LIMITE DELLA FREQUENZA RELATIVA IN UN NUMERO INFINITO DI PROVE
ES: SE SU 1000 LANCI DI UNA MONETA ESCE "TESTA" 520 VOLTE,
$P(\text{TESTA}) \approx 0{,}52$ (TIPO L'ACCURACY)

c) INTERPRETAZIONE BAYESIANA

- LA PROBABILITÀ È UNA MISURA SOGGETTIVA DEL GRADO DI FIDUCIA (O "CREDENZA") CHE UN INDIVIDUO HA NEL VERIFICARSI DI UN EVENTO, BASATA SU

CONOSCENZE PRECEDENTI (PRIOR) E AGGIORNATA CON NUOVE EVIDENZE (LIKELIHOOD)

- **Formula chiave**: Teorema di Bayes, permette di aggiornare la probabilità

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

$P(A)$: è la probabilità iniziale di A (prior)

$P(A|B)$: Probabilità aggiornata (posterior) di A dopo aver osservato B

Es → Scenario: Un paziente ha sintomi compatibili con una malattia rara (M).

Prior: $P(M) = 1\%$ (prevalenza della malattia nella popolazione)

Evidenza: Un test è positivo (T^+) con:

Sensibilità $P(T^+/M) = 95\%$ Probabilità di Test Positivo se il paziente è malato (i pazienti malati nel 95% dei casi sono risultati essere positivi a questo test)

Falsi Positivi $P(T^+/\neg M) = 10\%$ (positivo al test ma non malato)

Calcolo della Posterior

$$P(M/T^+) = \frac{P(T^+/M)\, P(M)}{P(T^+)} = 8{,}7\%$$

Nonostante il test positivo, la probabilità di essere malati è solo l'8,7% perché la malattia è rara.

Assiomi

Gli assiomi della probabilità sono come le regole del gioco che ci permettono di lavorare con l'incertezza in modo logico e senza contraddizioni senza ritrovarsi in casistiche assurde tipo probabilità negative o maggiori di 1

① Spazio degli Eventi possibili

- Ω (insieme degli esiti possibili)

Nel lancio di un dado $\Omega = \{1, 2, 3, 4, 5, 6\}$

- 6-Algebra (A): è una famiglia speciale di sottoinsiemi di un insieme

UNIVERSO Ω CHE SODDISFA 3 PROPIETÀ FONDAMENTALI. PENSA A Ω COME A TUTTI I POSSIBILI RISULTATI DI UN ESPERIMENTO (ES: I 6 NUMERI DI UN DADO), E ALLA σ-ALGEBRA COME AD UNA COLLEZIONE DI EVENTI SU CUI POSSIAMO CALCOLARE PROBABILITÀ IN MODO COERENTE.

a) CHIUSURA SOTTO COMPLEMENTO

SE UN EVENTO A È NELLA σ-ALGEBRA ALLORA ANCHE IL SUO OPPOSTO A^c (TUTTO CIÒ CHE NON È A) DEVE ESSERCI

ES: SE $A = \{ESCE\ PARI\} = \{2,4,6\}$ È INCLUSO ALLORA $A^c = \{ESCE\ DISPARI\} = \{1,3,5\}$ DEVE ESSERE INCLUSO.

b) CONTIENE Ω

L'EVENTO CERTO ("SUCCEDE QUALUNQUE COSA") DEVE ESSERE SEMPRE PRESENTE.

ES: PER UN DADO $\Omega = \{1,2,3,4,5,6\}$ DEVE ESSERE SEMPRE INCLUSO

c) CHIUSURA SOTTO UNIONI NUMERABILI

- SE HAI UNA SEQUENZA DI EVENTI $A_1, A_2, \dots$ NELLA σ-ALGEBRA, ALLORA LA LORO UNIONE $\bigcup_{n=1}^{\infty} A_n$ DEVE ESSERE NELLA STESSA ALGEBRA

ES: SE $A_1 = \{1\}, A_2 = \{2\}, \dots$ SONO INCLUSI, ALLORA $\bigcup_{n=1}^{6} A_n$ DEVE ESSERE INCLUSO

DA QUESTE SI RICAVANO ALTRE 2 CONSEGUENZE OVVERO:

i) L'INSIEME VUOTO È INCLUSO

↳ SE $\emptyset = \Omega^c$, Ω È INCLUSO x LA PROPIETÀ b

L'EVENTO "IMPOSSIBILE" (ES ESCE 7 QUANDO LANCI UN DADO) HA SEMPRE PROBABILITÀ 0

ii) INTERSEZIONI NUMERABILI

SE $A_1, A_2 \dots$ SONO NELL' σ-ALGEBRA ALLORA $\bigcap_{n=1}^{\infty} A_n$ (LA LORO INTERSEZIONE LO È ANCHE)

CONSIDERIAMO LA PROBABILITÀ COME UNA MISURA DI CERTEZZA CHE ASSEGNI AGLI EVENTI. KOLMOGOROV HA STABILITO 2 REGOLE D'ORO CHE OGNI FUNZIONE DI PROBABILITÀ DEVE SEGUIRE.

1. NORMALIZZAZIONE (REGOLA DEL 100%)

L'EVENTO CERTO Ω (TUTTI I POSSIBILI RISULTATI) HA PROBABILITÀ 1

ES: NEL LANCIO DI UN DADO $P(\{1,2,3,4,5,6\}) = 1$

PERCHÉ SIAMO SICURI CHE USCIRÀ UN NUMERO TRA 1 e 6

2. ADDITIVITÀ NUMERABILE (REGOLA DELLA SOMMA)

SE 2 O PIÙ EVENTI SONO INCOMPATIBILI (NON POSSONO VERIFICARSI ENTRAMBI) LA PROBABILITÀ DELLA LORO UNIONE È LA SOMMA DELLE SINGOLE PROBABILITÀ.

ES: PER UN DADO EQUO

$P(\{1\} \cup \{2\}) = P(1) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{i=1}^{\infty} P(A_n)$$

LA FUNZIONE DI PROBABILITÀ $P: A \to \mathbb{R}^+$ È DEFINITA SULL'INSIEME DI σ-ALGEBRA

LO IMPLICA DI AVERE UN UNIVERSO DI POSSIBILI RISULTATI (Ω) E UNA STRUTTURA DI EVENTI ATTRATTI (σ-ALGEBRA) CHE CONTIENE:

- TUTTI GLI STRUMENTI FONDAMENTALI
- PER OGNI STRUMENTO IL SUO OPPOSTO
- LA CAPACITÀ DI COMBINARE INFINITI STRUMENTI

IL TRIPLETTO (Ω, A, P) È CHIAMATO PROBABILITY SPACE

ESEMPIO (METEO)

- INSIEME DEI POSSIBILI RISULTATI: $\{$SOLE, PIOGGIA, NEVE$\} = \Omega$
- σ ALGEBRA, EVENTI A CUI ASSEGNARE PROBABILITÀ $\{\emptyset, \{$SOLE$\}, \{$PIOGGIA, NEVE$\}, \Omega\} = A$

FUNZIONE DI PROBABILITÀ CHE RISPETTA GLI ASSIOMI $= P: A \to \mathbb{R}^+$

TALE SPAZIO EVITA INCOERENZE POSSIBILI E GARANTISCE CHE TUTTI GLI EVENTI RILEVANTI SIANO MISURABILI

PROPRIETÀ DERIVANTI DAGLI ASSIOMI

- PROBABILITÀ DELL'EVENTO IMPOSSIBILE

$$P(\emptyset) = 0$$

DIM $\Omega = \Omega \cup \emptyset \cup \emptyset \ldots$ x ADDITIVITÀ $1 = 1 + P(0) + \ldots P(0) \Rightarrow P(0) = 0$

- MONOTONICITÀ

SE $A \subseteq B$ ALLORA $P(A) \leq P(B)$

ES $\{$ESCE 1$\} \subseteq \{$ESCE 1 o 2$\} \Rightarrow \frac{1}{6} \leq \frac{2}{6}$

- PROBABILITÀ DEL COMPLEMENTO

$$P(A^c) = 1 - P(A)$$

- REGOLA DELL'UNIONE

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

ESEMPIO: $A = \{PARI\} = \{2,4,6\}$, $B = \{$MAGGIORE DI 3$\} = \{4,5,6\}$

$$P(A \cup B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6}$$

$P(A) + P(B)$ CONTA 2 VOLTE L'INTERSEZIONE $A \cap B$, SOTTRAENDO $P(A \cap B)$ SI CORREGGE IL DOPPIO CONTEGGIO

ES: $A = \{$STUDENTI CHE AMANO LA MATEMATICA$\}$
$B = \{$STUDENTI CHE AMANO LA FISICA$\}$
$A \cap B = \{$STUDENTI CHE AMANO ENTRAMBI$\}$
$P(A \cup B) = 0.40 + 0.30 - 0.15 = 0.55$

- PROBABILITÀ DELL'UNIONE ATTRAVERSO I COMPLEMENTARI

$$P\left(\bigcup_n A_n\right) = 1 - P\left(\bigcap_n A_n^c\right)$$

- L'UNIONE $\bigcup_n A_n$ = "ALMENO UN EVENTO A_n SI VERIFICA"
- IL COMPLEMENTARE $\bigcap_n A_n^c$ = "NESSUN A_n SI VERIFICA"

ES

$A_n = \{$IL SISTEMA SUBISCE UN ATTACCO NEL GIORNO $n\}$

$\bigcup_{n=1}^{365} A_n$ = "ALMENO UN ATTACCO NELL'ANNO"

$\bigcap_{n=1}^{365} A_n^c$ = "NESSUN ATTACCO NELL'ANNO"

SE $P(A_n^c) = 0,99$ (99% PERCENTO DI SICUREZZA GIORNALIERA (NON C'È ATTACCO AL GIORNO n))

$$P\left(\bigcup_{n=1}^{365} A_n\right) = 1 - (0,99)^{365} \simeq 0,974$$

- PROBABILITÀ PER SUCCESSIONI MONOTONE

1) SUCCESSIONE CRESCENTE $A_1 \subseteq A_2 \subseteq A_3 \ldots$
SE $A = \bigcup_{n=1}^{\infty} A_n$

$$P(A) = \lim_{n \to \infty} P(A_n)$$

A_n CONTIENE TUTTO, TUTTE LE COMBINAZIONI DI INTERESSE PRESENTI NELL' σ-ALGEBRA

A_n APPROSSIMA SEMPRE MEGLIO A.

PASSIAMO A PROPRIETÀ PIÙ COMPLESSE, UNA VOLTA DEFINITO COS'È LO SPAZIO DI PROBABILITÀ (Ω, A, P) E ALCUNE DELLE SUE PROPRIETÀ PRINCIPALI

PROBABILITÀ CONDIZIONATA

LA PROBABILITÀ CONDIZIONATA RISPONDE ALLA DOMANDA: QUALE È LA PROBABILITÀ CHE SI VERIFICHI A SAPENDO CHE B È GIÀ ACCADUTO

È LA PROBABILITÀ CHE QUALCOSA ACCADA SLEGATA DA QUALSIASI CONOSCENZA CHE ABBIAMO GIÀ SULL'EVENTO

GIÀ ACCADUTO SI È VERIFICATO

LO TOLGO DAL TOTALE DEGLI EVENTI

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

ES: LANCIAMO UN DADO, NOI NON SAPPIAMO IL RISULTATO MA QUALCUNO CI DICE CHE NON SARÀ MAI 1

EVENTO A $p(\text{NUMERO} < 4)$ $A = \{1, 2, 3\}$

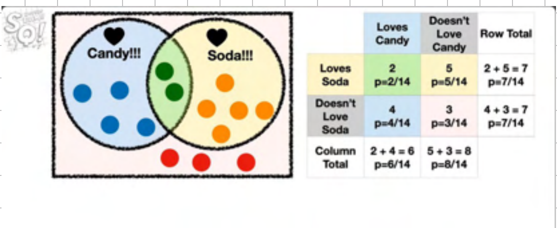
EVENTO B - SAPPIAMO GIÀ CHE IL NUMERO È > 1 $B = \{2, 3, 4, 5, 6\}$

$P(A/B) =$ INTUITIVAMENTE $= \frac{2}{5} = \frac{\text{2 IN QUANTO DUE USCITE (2 e 3)}}{\text{TOTALE DEI NUMERI CHE POSSONO USCIRE SAPENDO CHE 1 NON ESCE } (6-1) = 5} =$

$$= \frac{P(A \cap B)}{P(B)} = \frac{P(\{1,2,3\} \cap \{2,3,4,5,6\})}{P(\{2,3,4,5,6\})} = \frac{P(\{2,3\})}{\frac{5}{6}} = = \frac{\frac{2}{6}}{\frac{5}{6}} = \frac{2}{5}$$

TEOREMA DI BAYES

LA FORMULA DI BAYES COLLEGA $P(A/B)$ E $P(B/A)$, È UTILE IN QUANTO CI PERMETTE DI FARE APPROSSIMAZIONI SULLA PROBABILITÀ DI DETERMINATI EVENTI E RICAVARNE VALORE (L'APPROSSIMAZIONE VIENE FATTA X GRANDI DATASET) ED È ALLA BASE DELL'INTERPRETAZIONE BAYESIANA DELLA PROBABILITÀ)



	Loves Candy	Doesn't Love Candy	Row Total
Loves Soda	2 p=2/14	5 p=5/14	2+5=7 p=7/14
Doesn't Love Soda	4 p=4/14	3 p=3/14	4+3=7 p=7/14
Column Total	2+4=6 p=6/14	5+3=8 p=8/14	

SE ABBIAMO TUTTI I DATI È FACILE, SE NON SONO SICURO?

ES $p(\text{loves}) \approx 0{,}6$ MA NON NE SO SICURO

GRAZIE A BAYES POSSO CALCOLARE + ROBE.

$A = \{$ NON GLI PIACCIONO LE CARAMELLE $\}$ EVENTI

$B = \{$ GLI PIACE LA SODA $\}$

NON GLI PIACCIONO LE CARAMELLE MA LA SODA

VERE ENTRAMBE

PROBABILITÀ CONDIZIONATA = $P(A|B)$ = ASSODATO CHE GLI PIACE LA SODA QUAL È LA PROBABILITÀ CHE GLI PIACCIONO LE CARAMELLE? $= \dfrac{P(A \cap B)}{P(B)}$

$P(B|A)$ = ASSODATO CHE GLI PIACCIONO LE CARAMELLE QUAL È LA PROB CHE GLI PIACCIA LA SODA? $= \dfrac{P(B \cap A)}{P(A)}$

$$P(A \cap B) = P(B \cap A) = P(B)\, P(A|B) = P(B|A)\, P(A)$$

$$P(B|A) = \frac{P(B)\, P(A|B)}{P(A)}$$

$$P(A|B) = \frac{P(A)\, P(B|A)}{P(B)}$$

NOTA: EVENTI SONO INDIPENDENTI SE

$P(A \cap B) = P(A)\, P(B)$ IL VERIFICARSI DI UNO NON INFLUENZA LA PROBABILITÀ DELL'ALTRO

DISGIUNTI $\neq$ DA INDIPENDENTI

DISGIUNTI $\Rightarrow$ $A \cap B = \emptyset$ [ES LANCIO UN DADO, A = ESCE 1 B = ESCE 2, NON PUÒ USCIRE SIA UNO CHE 2 INSIEME]

$P(A \cap B) = 0$ MA NON È COSÌ

$P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{6} \neq 0$, POICHÉ $P(A \cap B) \neq P(A)\, P(B)$

I 2 EVENTI SONO DIPENDENTI PERCHÉ $P(A) + P(B) > 1$

SE A E B SONO DISGIUNTI $P(A|B) = 0$ SE ACCADE B A NON PUÒ ACCADERE

RANDOM VARIABLE

LE RV SONO MODI DI MAPPARE GLI OUTCOME DI RANDOM PROCESS IN NUMERI

RANDOM PROCESS (LANCIA UN DADO...)

Outcomes → Number   STO QUANTIFICANDO UN RISULTATO

TUTTI I RISULTATI POSSIBILI DI UN EVENTO

$$X : \Omega \to \mathbb{R}$$

↳ MAPPA OGNI ESITO w A UN NUMERO REALE

Es:

$$X = \begin{cases} 1 & \text{if Testa} \\ 0 & \text{if Croce} \end{cases}$$

(0, 1 SONO CASUALI POTREMMO ASSUMERE VALORI PARI A 100, 709...)

Y = LA SOMMA DELLE FACCE DI 7 DADI LANCIATI

STO QUANTIFICANDO L'OUTCOME DI UN RANDOM PROCESS

QUANTIFICARE GLI OUTCOMES MI PERMETTE DI USARE LA MATEMATICA, INVECE DI FARE RAGIONAMENTI LOGICI SU PIÙ DATI

ES P(sum of FACCE DOPO AVER LANCIATO I DADI è ≤ 30) = CAPISCI ...

$P(Y \leq 30)$ OPPURE $P(Y \text{ even})$

LE RV SONO $\neq$ DALLE VARIABILI NORMALI, NO HA SENSO $x+3=6$.... X PUÒ ASSUMERE MOLTI VALORI DIFFERENTI CON UNA DATA PROBABILITÀ → HA + SENSO DI PARLARE DI PROBABILITÀ CHE LA RV SIA UGUALE AD UN VALORE ....

$$X : \Omega \to \mathbb{R}$$

$$\forall x \in \mathbb{R} \Rightarrow \{ w \in \Omega : X(w) \leq x \}$$

QUESTO SIGNIFICA CHE L'OUTPUT DATO DALLA TRASFORMAZIONE DELLA RV DEVE SEMPRE ESSERE MAPPATO RISPETTO AD UN $w \in \Omega$

QUESTO EVENTO SI $\in \Omega$ MA PIÙ IN GENERALE $\in \mathcal{A}$ - σ-ALGEBRA OVVERO ALLE COMBINAZIONI DI NOSTRO INTERESSE DEL NOSTRO EVENTO RANDOMICO CHE STIAMO ANDANDO A MAPPARE NUMERICAMENTE TRAMITE L'UTILIZZO DELLA RV X

TRAMITE X POSSIAMO CALCOLARE $P(X \leq x)$ OVVERO LA PROBABILITÀ CHE X ASSUMA VALORI MINORI DI $x \in \mathbb{R}$. [x VA VERIFICATO È VALIDO E MAPPATO CORRETTAMENTE DA X]

Dato che X mappa l'insieme degli eventi possiamo trovare

Es: consideriamo X come la variabile random che mappa il lancio di 2 dati, la somma che deve usare corrisponde ai valori possibili in A

$$\{w : X(w) \le 4\} = \{(1,1),(1,2),(1,3),(2,1)(2,2),(3,1)\}$$

Dato che A include tutti i possibili sottoinsiemi di Ω (x noi sarà sempre così)

quindi l'insieme è valido

$$P(X \le 4) = \frac{6}{36} = \frac{1}{6}$$

Dato che X deve garantire gli $w \in A$ x definizione di R.V posso anche

- considerare $\{w : X(w) > a\}$ ovvero il complementare di $\{w : X(w) \le a\}$
- $\{w : a < X(w) \le b\}$ intersezione di $\{w : X(w) > a\}$ e $\{w : X(w) \le b\}$ (sono sempre contenuti in σ-algebra x la sua definizione)
- $\{w : X(w) = x\}$ intersezione infinita di intervalli stretti

$$= \bigcap_n \{w : x - \tfrac{1}{n} < X(w) \le w$$

valori possibili x $P(X = x)$

Perché intersezione? Intersecando tutti questi insiemi $x(n \to \infty)$ restringiamo l'intervallo fino a includere solo i risultati dove $X(w) = X$ esattamente. È come zoomare sempre + escludendo i valori superflui

Funzione di ripartizione (CDF)

La funzione CDF è una funzione $F_X(x)$ che risponde alla domanda, qual è la probabilità che X assuma un valore minore o uguale a x

$$F_X(x) = P(X \le x)$$

La CDF ha le seguenti proprietà

① Limitazione: $0 \le F_X(x) \le 1 \ \forall x \in \mathbb{R}$, questa proprietà deriva dal concetto di probabilità, dalla funzione P che compone il nostro spazio vettoriale [la probabilità può valere 100% o 0%)

② Monotonia: Se $x_1 < x_2$ allora $F_X(x_1) \le F_X(x_2)$ la FDC è monotona decrescente. Intuitivamente, se consideriamo un valore più grande di x_2, stiamo includendo tutti gli eventi già considerati in x_2 più potenzialmente

ALTRI EVENTI, QUINDI LA PROBABILITÀ NON PUÒ DIMINUIRE

③ LIMITI ASINTOTICI: $\lim_{x \to -\infty} F_X(x) = 0$ e $\lim_{x \to \infty} F_X(x) = 1$, IL PRIMO DESCRIVE UN EVENTO QUASI IMPOSSIBILE (L'INSIEME VUOTO $\varnothing$ CHE FA PARTE A) IL SECONDO UN EVENTO QUASI CERTO

④ CONTINUITÀ A DESTRA: $\lim_{x \to x_0^+} F_X(x) = F_X(x_0)$ LA FDR È CONTINUA A DESTRA IN OGNI PUNTO. CIÒ SIGNIFICA CHE SE CI AVVICINIAMO A x_0 DA VALORI LEGGERMENTE MAGGIORI IL LIMITE DELLA FUNZIONE COINCIDE CON IL VALORE DELLA FUNZIONE IN x_0

CONSEGUENZE IMPORTANTI:

i) PROBABILITÀ DI INTERVALLI: $P(a < X \leq b) = F_X(b) - F_X(a)$ [PROB CHE X CADA IN UN INTERVALLO SEMIAPERTO

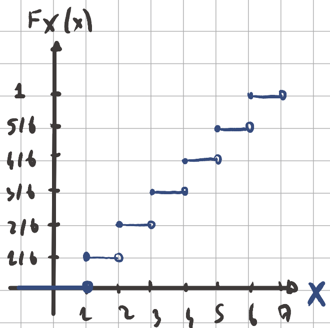
ii) PROBABILITÀ DI PUNTI SINGOLI: $P(X = x_0) = F_X(x_0) - \lim_{x \to x_0^-} F_X(x)$

$F_X(x_0)$: $P(X \leq x_0)$

$\lim_{x \to x_0^-} F_X(x)$: PROBABILITÀ CHE X SIA STRETTAMENTE MINORE DI x_0 $P(X < x_0)$

ES CONSIDERO LA VARIABILE ALEATORIA X CHE MI DESCRIVE IL LANCIO DI UN DADO

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1/6 & \text{if } 1 \leq x < 2 \\ 2/6 & \text{if } 2 \leq x < 3 \\ 3/6 & \text{if } 3 \leq x < 4 \\ 4/6 & \text{if } 4 \leq x < 5 \\ 5/6 & \text{if } 5 \leq x < 6 \\ 1 & \text{if } x \geq 6 \end{cases}$$



$P(X = 3) = F(3) - F(3^-)$

$= 3/6 - \lim_{x \to 3^-} F(x) = \frac{3}{6} - \frac{2}{6} = \frac{1}{6}$

Discrete random variable

Una variabile aleatoria X è definita discreta se assume un numero finito o un'infinità di valori numerabili. Ciò le distingue da quelle continue che possono assumere qualsiasi valore in un intervallo.

Per una r.v discreta X, possiamo definire la funzione di massa di probabilità f.m.p o densità discreta

$$f_X(x) = P(X=x)$$ [ha senso solo con r.v discrete]

Questa funzione assegna a ogni possibile valore di X la sua probabilità di occorrenza.

Proprietà

① Non negativa $f_X(x) \geq 0$ x ogni x. Le probabilità non possono mai essere negative

② Sparsità $f_X(x) = 0$ per tutti i valori di x tranne al massimo un infinità numerabile di valori (quelli assunti da X).

③ Normalizzazione: $\sum_{x \in S} f_X(x) = 1$, dove S è il supporto di X ovvero l'insieme dei valori per cui $f_X(x) > 0$

Se conosciamo $f_X(x)$ ovvero la funzione di densità discreta è possibile a sua volta calcolare $F_X(x) = \sum_{t \leq x} f_X(t) = P(X \leq x)$

Ciò ci dice come è chiaro che sia che $P(X \leq x)$ è pari alla somma delle probabilità di tutti i possibili valori di X minori o uguali a x. La sommatoria è estesa a tutti i valori di t nel supporto di X ovvero $\in S$ t.c $t \leq x$ [ovviamente gli altri non hanno contributo dato che avendo era valore 0]

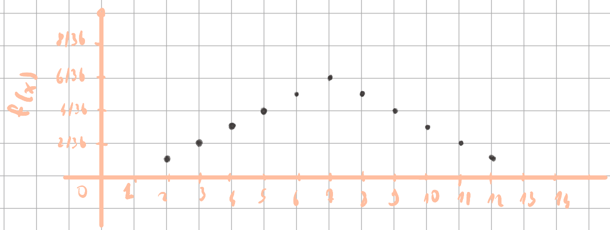
Es: somma dei valori ottenuti lanciando 2 dadi

S = supporto = $\{2,3,4,5,6,7,8,9,10,11,12\}$

$p(X=2) = P(1,1) = \frac{1}{36}$

$p(X=3) = P(1,2) + P(2,1) = \frac{2}{36}$

⋮



FORMA A CAMPANA, CI SONO INFATTI PIÙ COMBINAZIONI CHE DANNO 7

CONOSCERE LA f.m.p EQUIVALE A CONOSCERE TUTTE LE PROBABILITÀ ASSOCIATE AD UNA VARIABILE ALEATORIA (R.V)

RV DISCRETE CON SUPPORTO INFINITO

FINO AD ORA ABBIAMO CONSIDERATO RV CON SUPPORTO FINITO, COME IL LANCIO DI UN DADO. MA COSA SUCCEDE SE IL SUPPORTO DIVENTA INFINITO NUMERABILE?? OVVERO CONTIENE UN INFINITÀ DI VALORI CHE POSSONO ESSERE MESSI IN CORRISPONDENZA BIUNIVOCA UNO RISPETTO ALL'ALTRO.

ES: LANCIO DI UNA MONETA N VOLTE (SUPPORTO FINITO)

LO SPAZIO CAMPIONARIO CONTIENE TUTTE LE POSSIBILI SEQUENZE DI TESTE H e CROCI C DI DI LUNGHEZZA K

PER N=3 LO SPAZIO CAMPIONARIO È

$$\Omega = \{(HHH), (HHT), (HTH), (HTT), (THH), (THT), (TTT)\}$$

SE LA MONETA È EQUA OGNI SEZIONE HA LA STESSA PROBABILITÀ DI VERIFICARSI CIOÈ $p = \left(\frac{1}{2}\right)^{n=3} = \frac{1}{8}$

LO SPAZIO È FINITO COME IL SUPPORTO HA 2^n ELEMENTI

ES: LANCIO DI UNA MONETA FINO ALLA PRIMA TESTA (SUPPORTO INFINITO)

QUESTO ESPERIMENTO PUÒ POTENZIALMENTE CONTINUARE ALL'INFINITO MA SI FERMERÀ PRIMA O POI CON CERTEZZA A MENO CHE LA MONETA NON ABBIA PROBABILITÀ 0 DI FAR USCIRE TESTA.

LO SPAZIO CAMPIONATO

$$\Omega = \{(H), (T,H), (T,T,H), \dots\}$$

QUESTO SPAZIO È CAMPIONATO INFINITO NUMERABILE, POICHÉ LE SEQUENZE POSSIBILI SONO IN CORRISPONDENZA CON I NUMERI NATURALI {SEQUENZA CON n CROCI SEGUITE DA UNA TESTA CORRISPONDE AL NUMERO n}

CALCOLO DELLA PROBABILITÀ

- CONSIDERANDO L'EVENTO A = "LA PRIMA MONETA È TESTA" $= |(H)| \Rightarrow P(A) = \frac{1}{2}$ SE LA MONETA È EQUA

IL COMPLEMENTARE DI A È A^c = "LA PRIMA MONETA È CROCE" $= |(TH), (TTH), \dots|$, x LA PROPRIETÀ DELLA COMPLEMENTARIETÀ $P(A^c) = 1 - P(A) = \frac{1}{2}$

- CONSIDERANDO L'EVENTO B = "LA SECONDA MONETA È TESTA" $= |(TH)|$ DATO CHE I LANCI SONO INDIPENDENTI LA PROBABILITÀ CHE AL SECONDO LANCIO SI OTTENGA ANCORA TESTA È $1/2$

$$P(B) = P(\text{PRIMO LANCIO} = T) \times P(\text{SECONDO LANCIO} = H) = 1/2 \times 1/2 = 1/4$$

$$P(B^c) = P(\text{PRIMO LANCIO} = T) \times P(\text{PRIMO LANCIO} = T) = 1/4$$

Proseguendo x una qualunque sequenza w, $P(\{w\}) = \frac{1}{2}^n$ n è la lunghezza della moneta

NOTA: la somma di probabilità si usa quando si vuole trovare la probabilità che si verifichi almeno uno tra i diversi eventi (unione di eventi)
il prodotto di probabilità si usa quando si vuole trovare la probabilità che si verifichino tutti gli eventi in una specifica sequenza

Generalizzazione per moneta non equa

Se la probabilità di ottenere testa è p e quindi di ottenere croce $(1-p)$, la probabilità di ogni sequenza diventa
$$p(\{w\}) = p(1-p)^{n-1}$$, n è la lunghezza della sequenza w

Distribuzione geometrica

x descrivere l'evento aleatorio

Se definiamo una variabile aleatoria X che associa ad ogni esito il numero di croci necessarie prima di ottenere una testa, otteniamo:

$$X(\{H\}) = 0,\ X(\{T,H\}) = 1,\ X(\{T,T,H\}) = 2 \ldots$$

Il supporto di X è l'insieme dei numeri naturali positivi N^+
La funzione di massa di probabilità che descrive la funzione geometrica è

$$f_X(x) = \begin{cases} p(1-p)^x & x \in N^+ \\ 0 & \text{altrimenti} \end{cases}$$

Questa distribuzione modella il numero di fallimenti (croci) prima di ottenere croce. Ha memoria nulla: la probabilità di ottenere testa al prossimo lancio non dipende dai lanci precedenti.

Continuous random variable

Sono variabili che possono assumere qualsiasi valore in R o in un suo sottoinsieme.
La c.d.f funzione o distribuzione cumulativa di una r.v X $\Rightarrow F_X(x) = P(X \le x)$

Essendo X continua, lo sarà anche la sua c.d.f (dato che $X \in R$ anche $x \in R$ non c'è alcun supporto). La principale conseguenza di ciò è che la probabilità di una variabile X di assumere esattamente un valore specifico è = a 0 [se ci pensi dato che x può assumere un infinità di valori è logico pensare che sia così] $P(X=x)=0$

$$\lim_{x \to x_0^-} F_X(x) = \lim_{x \to x_0^+} F_X(x) = F_X(x_0)$$

$$\left.\begin{array}{l} \\ \text{Ricorda} \quad P(X = x_0) = F_X(x_0) - \lim_{x \to x_0^-} F_X(x) \end{array}\right\} \Rightarrow P(X = x_0) = 0$$

Una conseguenza di ciò è che x le variabili aleatorie continue

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b)$$

Aggiungere o rimuovere i punti estremi non modifica la probabilità dato che la probabilità di ogni singolo punto è 0

Come nelle R.V discrete abbiamo definito la funzione di massa, lo facciamo anche qua introducendo la probability density

PROBABILITY DENSITY

$f: \mathbb{R} \to \mathbb{R}$ è una densità se e solo se

1. $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
2. f is integrable over $\mathbb{R}$
3. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

A questo punto data una R.V X se esiste una funzione densità f t.c

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

Allora f_X è la funzione di densità di probabilità della R.V X
(Non se $F_X(x) = P(X \leq x)$ - allora l'area sottostante alla funzione  che descrive la variabile aleatoria continua non è altro che la probabilità cercata.)

È importante notare che la f.d.p non è unica questo perché la probabilità si calcola tramite integrali, e cambiare il valore della funzione in un insieme di misura nulla non modifica i valori dell'integrale. f_X of X

A sua volta se F_X è differenziabile, allora

$$f_X(x) = \frac{d\,F_X(x)}{dx}$$

INTERPRETAZIONE

LA FUNZIONE DI DENSITÀ DI PROBABILITÀ NON È UNA DESCRIZIONE DEL PROCESSO CASUALE IN SÉ QUANTO UN MODELLO MATEMATICO CHE CATTURA IL COMPORTAMENTO STATISTICO DEL PROCESSO SU LARGA SCALA O IN TERMINI DI PROBABILITÀ

IMMAGINA DI MISURARE L'ALTEZZA ESATTA DELLE PERSONE. L'ALTEZZA È UNA VARIABILE CONTINUA PUÒ ESSERE 170 cm 170,1 cm, 170,01 cm 170,001 cm... LA PROBABILITÀ CHE QUALCUNO SIA ALTO ESATTAMENTE 170,00.... cm È TECNICAMENTE ZERO PERCHÉ CI SONO INFINITI VALORI POSSIBILI

SAPPIAMO PERÒ CHE MOLTE PERSONE HANNO UN ALTEZZA INTORNO A 170 cm, COME SI PUÒ ESPRIMERE QUESTA IDEA DI CONCENTRAZIONE DI PROBABILITÀ MATEMATICAMENTE? → TRAMITE LA FUNZIONE DENSITÀ DI PROBABILITÀ!!

IL VALORE $f(x)$ IN UN PUNTO NON È ESATTAMENTE LA PROBABILITÀ MA RAPPRESENTA QUANTO È DENSA O CONCENTRATA LA PROBABILITÀ INTORNO A QUEL PUNTO

SE $f(170) = 0,05$ E $f(190) = 0,01$, SIGNIFICA CHE C'È UNA MAGGIORE CONCENTRAZIONE DI PROBABILITÀ INTORNO A 170 cm RISPETTO A 190 cm

SE VOGLIAMO CALCOLARE LA PROBABILITÀ DI UN INTERVALLO INFINITESIMALE $P(x \leq X \leq x + dx) = f(x)dx$

LA MASSA DI UNA PICCOLA SEZIONE DI MATERIALE È DENSITÀ × VOLUME → LA MASSA È LA NOSTRA PROBABILITÀ + C'È MASSA PIÙ È PROBABILE CHE ACCADA

× INTERVALLI PIÙ GRANDI SI SOMMANO TUTTI QUESTI CONTRIBUTI INFINITESIMALI

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

PIÙ IN GENERALE SE $X \in A$ È UN EVENTO, POSSIAMO CALCOLARE LA SUA PROBABILITÀ COME $P(X \in A) = \int_A f_X(x)\,dx$

QUINDI IL VALORE $f_X(x_0)$ NON RAPPRESENTA LA PROBABILITÀ CHE $X = x_0$ (CHE È 0), MA FORNISCE UN INDICAZIONE DI QUANTO SIA DENSA LA PROBABILITÀ INTORNO A x_0

RANDOM VECTOR

UN RANDOM VECTOR m-DIMENSIONAL È UNA GENERALIZZAZIONE DEL CONCETTO DI VARIABILE ALEATORIA A PIÙ DIMENSIONI. È FONDAMENTALMENTE UN VETTORE $X = (X_1, X_2, X_3 \ldots X_m)$ LE CUI COMPONENTI SONO TUTTE VARIABILI RANDOMICHE. CIÒ CI PERMETTE DI MODELLARE SITUAZIONI IN CUI SONO PRESENTI PIÙ QUANTITÀ INCERTE CHE POSSONO ESSERE CORRELATE TRA LORO.

Per esempio se studiamo la salute delle persone, potremmo considerare un vettore aleatorio bidimensionale (X,Y) dove X rappresenta il peso e Y l'altezza.

Cumulative Joint Distribution Function

La ---- di un R.V X è data da $F_X(x) = P(X_1 \le x_1, \dots X_m \le x_m)$

È l'unione di tutte le cdf delle variabili aleatorie

Questa funzione ci dà la probabilità che ciascuna componente del vettore aleatorio sia simultaneamente minore o uguale al corrispondente valore specificato nel vettore $x=(x_1, x_2, \dots x_m)$

L'evento è congiunto: tutte le disuguaglianze devono essere soddisfatte contemporaneamente.

Esempio x il caso binomiale $(X,Y) \Rightarrow F_{X,Y}(x,y) = P(X \le a, Y \le b)$, probabilità che X sia minore o uguale ad a e $Y \le$ a b

Distribuzioni Marginali

Quando abbiamo random variabile che consideriamo insieme la distribuzione marginale di una di esse ci dice come quella specifica variabile si distribuisce quando ignoriamo o integriamo via le altre.

ES: immaginiamo di avere informazioni sul genere e sul colore di capelli di un gruppo di persone.

	Cap Neri	Cap Castani	Cap Biondi	Totale (Marginale)
Maschi	20	15	5	40
Femmine	25	20	15	60
Totale (marginale)	45	35	20	100

I numeri all'interno rappresentano le frequenze congiunte: quante persone hanno sia la caratteristica della riga che della colonna

I totali di riga e colonna sono le <u>frequenze marginali</u>: tipo 40 persone sono maschi, indipendentemente dal colore dei capelli

Formule

Random Vector Discreti

- Funzione di massa di probabilità congiunta per un random vector discreto X.
  Rappresenta la probabilità che X assuma esattamente il valore x

$$f_X(x) = P(X=x) = P(X_1=x_1, X_2=x_2 \dots X_m=x_m)$$

Ci permette di esprimere la probabilità che tutte le componenti assumano

SIMULTANEAMENTE DETERMINATI VALORI → PROBABILITÀ DELL'INTERSEZIONE DI m EVENTI ELEMENTARI

ES: SE $X=(X_1,X_2)$ E ESSI RAPPRESENTANO I RISULTATI DI 2 DADI ALLORA $fX(3,5)=P(X_1=3, X_2=5)=1/36$

- **FUNZIONE DI DISTRIBUZIONE CUMULATIVA DISCRETA**

$$FX(x)=\sum_{y_1<x_1}\cdots\sum_{y_m<x_m} fX(y)$$

RAPPRESENTA LA PROBABILITÀ CHE CIASCUNA COMPONENTE SIA $\leq$ DEL CORRISPONDENTE VALORE IN X.
LA FDC SOMMA LE PROBABILITÀ DI TUTTI I POSSIBILI RISULTATI CHE SODDISFANO LA CONDIZIONE $X\leq x$. IN UN CONTESTO MULTIDIMENSIONALE DOBBIAMO SOMMARE LE PROBABILITÀ DI TUTTI I PUNTI y t.c. OGNI COMPONENTE DI y è $\leq$ ALLA RISPETTIVA COMPONENTE DI x

y è UNA VARIABILE MUTA, RAPPRESENTA TUTTI I POSSIBILI VALORI CHE IL RANDOM VECTOR X PUÒ ASSUMERE.

ES DADO DOPPIO $X=(X_1,X_2)$ $FX(x_1,x_2)$ è LA PROB CHE IL PRIMO DADO DIA UN RISULTATO MINORE DI x_1 E IL SECONDO DI x_2

$$Fx(x_1,x_2)=\sum_{y_1<x_1}\sum_{y_2<x_2} f_X(y)$$ → SIGNIFICA SOMMA LE PROBABILITÀ $fX(y_1,y_2)$ X TUTTE LE POSSIBILI COMBINAZIONI DI y_1, y_2 t.c. $y_1\leq x_1$ e $y_2\leq x_2$

$$Fx(3,2)=fx(1,1)+f_X(1,2)+fx(2,1)+\ldots=\frac{1}{36}+\frac{2}{36}\cdots=\frac{1}{6}$$

È COME SOMMARE TUTTI I PUNTI DISCRETI FACENTI PARTE DI UN IPER RETTANGOLO

RANDOM VECTOR CONTINUI

- **FUNZIONE DI DENSITÀ DI PROBABILITÀ CONGIUNTA** $X.length=m$

$$fX(x)=\frac{\partial^m}{\partial x_1\partial x_2\cdots\partial_m}FX(x)$$

È LA DERIVATA PARZIALE MISTA DI ORDINE m DELLA FUNZIONE DI DISTRIBUZIONE CUMULATIVA

→ LA DENSITÀ IN UN PUNTO X RAPPRESENTA IL TASSO DI ACCUMULO DI PROBABILITÀ INTORNO A QUEL PUNTO (DENSITÀ). MATEMATICAMENTE LA DERIVATA F.D.C RAPPRESENTA

DIGRESSIONE SULLE DERIVATE PARZIALI MISTE

QUANDO ABBIAMO UNA FUNZIONE DI PIÙ VARIABILI, COME d.c.f. POSSIAMO CALCOLARE LE DERIVATE RISPETTO A CIASCUNA DELLE VARIABILI SEPARATAMENTE.

DERIVATE PARZIALI

- **DERIVATA PARZIALE** SEMPLICE, SE ABBIAMO $F(x,y) \rightarrow \frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}$ — CI DICE COME CAMBIA F QUANDO y È COSTANTE E VICEVERSA

DERIVATA PARZIALE DI SECONDO ORDINE $\frac{\partial^2 F}{\partial x^2}$ o $\frac{\partial^2 F}{\partial x \partial y}$

DERIVATA PARZIALE MISTA: QUANDO DERIVIAMO RISPETTO A VARIABILI DIVERSE IN SEQUENZA ES $\frac{\partial^2 F}{\partial x \partial y}$

QUINDI LA DERIVATA PARZIALE MISTA DI ORDINE m SIGNIFICA DERIVARE UNA VOLTA RISPETTO A CIASCUNA DELLE m VARIABILI

$\frac{\partial^m F}{\partial x_1 \partial x_2 \dots \partial x_m}$ DERIVA F RISPETTO A x_1, POI DERIVA IL RISULTATO RISPETTO A x_2 ...

CIÒ È IMPORTANTE PERCHÉ SE LA FUNZIONE DI DISTRIBUZIONE CUMULATIVA $F(x_1, x_2, \dots x_m)$ CI DICE LA PROBABILITÀ CHE IL VETTORE ALEATORIO X CADA NELLA REGIONE $(-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_m]$.

LA DERIVATA PARZIALE MISTA MISURA IL TASSO DI ACCUMULO DI PROBABILITÀ IN UN PUNTO SPECIFICO CI DICE QUANTO VELOCEMENTE CRESCE LA PROBABILITÀ QUANDO CI MUOVIAMO SIMULTANEAMENTE IN TUTTE LE DIREZIONI.

· FUNZIONE DI DISTRIBUZIONE CUMULATIVA CONTINUA

$$F_X(x) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_m} f_X(y)\, dy_1\, dy_2 \dots dy_m$$

INTEGRALE MULTIPLO DELLA DENSITÀ. PER CALCOLARE LA PROBABILITÀ CHE X SIA IN UNA REGIONE DELLO SPAZIO INTEGRIAMO LA DENSITÀ SU QUELLA REGIONE. $P(X \le x)$ INTEGRALE DELLA DENSITÀ SULLA REGIONE ILLIMITATA $(-\infty, x_1] \times \dots \times (-\infty, x_m]$

$\hookrightarrow$ y È ANCHE QUI UN VETTORE $y = (y_1, y_2, \dots y_m)$ CHE RAPPRESENTA UN PUNTO GENERICO NELLO SPAZIO m-DIMENSIONALE. L'INTEGRALE MULTIPLO STA SOMMANDO (INTEGRANDO) I VALORI DELLA DENSITÀ $f_X(y)$ SU TUTTI I PUNTI y t.c. $y_1 \le x_1$, $y_2 \le x_2$ ...

DISTRIBUZIONI MARGINALI

QUESTA FORMULA CALCOLA LA DISTRIBUZIONE MARGINALE DISCRETA DELLA PRIMA COMPONENTE X_1 DI UN VETTORE ALEATORIO DISCRETO

$$f_{X_1}(x_1) = \sum_{y_2 \in S(X_2)} \sum_{y_3 \in S(X_3)} \cdots \sum_{y_m \in S(X_m)} f_X(x_1, y_2, \ldots, y_m)$$

QUESTA MI PERMETTE DI TROVARE LA PROBABILITÀ CHE $X_1 = x_1$ INDIPENDENTEMENTE DAI VALORI DELLE ALTRE COMPONENTI, POSSIAMO SOMMARE LE PROBABILITÀ DI TUTTI GLI EVENTI POSSIBILI IN CUI $X_1 = x_1$ E LE ALTRE COMPONENTI ASSUMONO QUALSIASI VALORE NEL LORO SUPPORTO

ES: LANCIO DI 2 DADI X TROVARE LA PROBABILITÀ CHE IL PRIMO DADO DIA 3, SOMMIAMO LE PROBABILITÀ DI TUTTI I POSSIBILI RISULTATI IN CUI IL PRIMO DADO È 3

$$f_{X_1}(3) = \sum_{y_2 \in \{1,2,3,4,5,6\}} f_{X_1}(3, y_2) = f_X(3,1) + f_X(3,2) + f_X(3,3) + \ldots f_X(3,6) = \frac{6}{36} = \frac{1}{6}$$

SUPPORTO

DISTRIBUZIONE MARGINALE CONTINUA

$$f_{X_1}(x_1) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_X(x_1, y_2, y_3, \ldots y_m)\, dy_2 \ldots dy_m$$

DISTRIBUZIONE MARGINALE DELLA PRIMA COMPONENTE X_1 DI UN VETTORE ALEATORIO CONTINUO
ANALOGAMENTE AL CASO DISCRETO X TROVARE LA DENSITÀ MARGINALE DI X_1 IN x_1 INTEGRIAMO LA DENSITÀ CONGIUNTA RISPETTO A TUTTE LE ALTRE VARIABILI.
(È COME SOMMARE)

INDIPENDENZA STATISTICA

$$X \perp\!\!\!\perp Y \iff F_{X,Y}(x,y) = F_X(x) F_Y(y)$$

L'INDIPENDENZA SIGNIFICA CHE CONOSCERE IL VALORE DI UNA VARIABILE NON FORNISCE ALCUNA INFORMAZIONE SUL VALORE DELL'ALTRA. IN TERMINI PROBABILISTICI, QUESTO SI TRADUCE NELLA FATTORIZZAZIONE DELLA PROBABILITÀ CONGIUNTA COME PRODOTTO DELLE PROBABILITÀ MARGINALI (PRODOTTO ALLORA TUTTE IN SEQUENZA)

PER LE VARIABILI DISCRETE

L'INDIPENDENZA EQUIVALE A $f_{XY}(x,y) = f_X(x) f_Y(y)$

$$\Updownarrow$$

$$P(X=x, X=y) = P(X=x) P(Y=y)$$

PER VARIABILI CONTINUE

L'INDIPENDENZA EQUIVALE A $f_{XY}(x,y) = f_X(x) f_Y(y) \iff X \perp\!\!\!\perp Y$

Densità Condizionata

Sapendo che una variabile aleatoria ha assunto un certo valore, come si distribuisce un'altra variabile aleatoria?

La densità condizionata ci permette di rispondere a ciò, dato X dato Y=y

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}, \quad \forall y \in S(Y)$$

S(Y) è il supporto di Y ovvero l'insieme di valori che Y può assumere con esito positivo

Nota Questa formula deriva dalla definizione di probabilità condizionale. Nel caso discreto sappiamo che

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

la densità congiunta $f_{X,Y}(x,y)$ rappresenta la probabilità congiunta
la densità marginale $f_Y(y)$ rappresenta la probabilità di Y=y

ES Sto studiando l'altezza X e il peso Y nelle persone. La densità congiunta $f_{X,Y}(x,y)$ ci dice quanto è probabile trovare una persona con altezza x e peso y

Se sappiamo che una persona ha peso 70 kg (Y=70) come si distribuisce la sua altezza??
La densità condizionata risponde a ciò $f_{X|Y}(x,70)$

Matematicamente, tagliamo la distribuzione congiunta lungo la linea Y=70, poi normalizziamo, dividendo x $f_Y(70)$ per la distribuzione di probabilità valida

Regola di Bayes

Ci permette di invertire la condizione, la formula è

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

Formula per "l'inversa" x trovare il comportamento di una variabile rispetto all'altra.

$$f_X(x) = \int f_{X,Y}(x,y)\,dy = \int f_{X|Y}(x,y)\, f_Y(y)\,dy$$

fissa stesso valore — fissa stesso valore — tutti le possibili y in $\mathbb{R}$

La densità marginale di X è la somma (integrale) delle densità congiunte su tutti i possibili valori di Y, che equivale alla somma ponderata delle densità condizionali, dove i pesi sono dati dalla densità marginale di Y

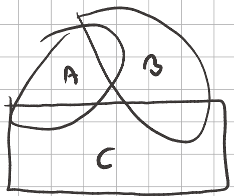
Indipendenza nella probabilità condizionata!!

X e Y sono condizionalmente indipendenti dato Z se, una volta noto il valore di Z, la conoscenza di X non fornisce ulteriori informazioni su Y e viceversa.

$$(X \perp\!\!\!\perp Y)\,|\,Z \iff F_{X,Y|Z}(x,y|z) = F_{X|Z}(x|z)\cdot F_{Y|Z}(y|z)$$

$$X \perp\!\!\!\perp Y \iff F_{X,Y}(x,y) = F_X(x)\,F_Y(y)$$

Sebbene 2 variabili possano essere dipendenti in generale, possono diventare indipendenti quando si conosce una terza variabile



Se sappiamo C, siamo in C, conoscere A non ci dà informazioni su B

Valore Atteso (Media)

Il valore atteso di una variabile aleatoria rappresenta il centro di gravità della sua distribuzione, il valore medio che otteniamo se ripetessimo l'esperimento infinite volte

Per una R.V discreta

$$E[X] = \sum_{x \in S} x \cdot f_X(x)$$

Media ponderata di tutti i possibili valori che X può assumere dove ciascun valore x viene pesato con la sua probabilità $f_X(x)$.
$f_X(x) = P(X=x)$. Questo è l'outcome medio della R.V discreta, seguendo il suo

SUPPORTO S [IL CUADINO MEDIO CONSIDERANDO LE PERDITE (PENSA A UNA BET)]

R.V CONTINUE

L'ASPETTAZIONE DI PRIMO ORDINE ESPRIME LO STESSO CONCETTO DI PRIMA SOLO CHE NEL CASO CONTINUO NON ABBIAMO $P(X=x)$ CHE È 0 CONSIDERIAMO LA DENSITÀ

$f_X(x)$ SE CONSIDERO UN RETTANGOLINO INFINITESIMALE dx → L'INTEGRALE MI PERMETTE DI SOMMARE TUTTI I RETTANGOLINI CHE RAPPRESENTANO LA MIA $P(x-\Delta x < X < x+\Delta x)$

NON CHE IL PESO DA DARE ALLA MIA INTEGRAZIONE / SOMMA)

$$E_X[X] = \int_S x f_X(x) dx$$

S È SEMPRE IL SUPPORTO OVVERO DOVE LA DENSITÀ NON ASSUME VALORE

(Varianza)

MISURA QUANTO I VALORI DI UNA VARIABILE ALEATORIA SONO DISPERSI ATTORNO AL SUO VALORE ATTESO (MEDIA). MEDIAMENTE QUANTO SONO DISTANTI I VALORI DAL VALORE ATTESO?! – È LETTERALMENTE CIÒ CHE CALCOLA LA FORMULA, OVVIAMENTE PESANDO IL TUTTO IN BASE ALLA PROBABILITÀ DI QUEL VALORE.

$$var(X) = E[(X-E[X])^2] = E[X^2]-E[X]$$
$$= E[X^2] - E[X]$$

DISCRETA R.V

$$E_X[X^2] = \sum_{x \in S} x^2 f_X(x)$$

CONTINUE R.V

$$E_X[X^2] = \int_S x^2 f_X(x) dx$$

L'ELEVAZIONE AL QUADRATO VIENE FATTA X PRATICITÀ ED INOLTRE DA PIÙ PESO AGLI OUTLIER

Proprietà

$var(X) \geq 0$

$var(X) = 0$

$var(cX) = c^2 var(X)$

$var(X+Y) = var(X) + var(Y) \Leftarrow X \perp Y$

Deviazione Standard

NON È ESATTAMENTE LA MEDIA DELLE DISTANZE DALLA MEDIA, MA FORNISCE UNA RAPPRESENTAZIONE QUANTITATIVA DI QUANTO I VALORI TIPICAMENTE SIANO DISTANTI DAL VALOR MEDIO

$\sigma = std(X) = \sqrt{Var(X)}$

CIÒ È UTILE PERCHÉ CI PERMETTE DI STIMARE QUANTI VALORI CADONO X DETERMINATE REGIONI ES: DISTRIBUZIONE NORMALE IL 68% DEI VALORI CADE ENTRO 1 DEVIAZIONE STANDARD DALLA MEDIA $(\mu \pm \sigma)$ ...

IN GENERALE IL MOMENTO N-ESIMO DI UNA R.V. X È DEFINITO COME $E_X[X^n]$

DISCRETO

$$E_X[X^n] = \sum_{x \in S} x^n f_X(x)$$

CONTINUO

$$E_X[X^n] = \int_S x^n f_X(x)\, dx$$

I MOMENTI SONO IMPORTANTI INDICATORI, CHE CARATTERIZZANO LA FORMA DI UNA DISTRIBUZIONE DI PROBABILITÀ

ES IL TERZO MOMENTO $E[X^3]$ È CORRELATO ALL'ASIMMETRIA DELLA FUNZIONE
$E[X^4]$ ALLA CURTOSI OVVERO QUANTO LA DISTRIBUZIONE È APPUNTITA

IN GENERALE I MOMENTI CENTRALI INVECE, MISURANO LA DISTRIBUZIONE RISPETTO ALLA MEDIA.

IL N-ESIMO MOMENTO CENTRALE È DEFINITO COME $E[(X - E[X])^n]$

ES: IL 3 MOMENTO, MISURA L'ASIMMETRIA DELLA DISTRIBUZIONE, SE È POSITIVO HA UNA CODA PIÙ LUNGA VERSO DESTRA, IL CONTRARIO SE È NEGATIVO.

IL VALORE ATTESO PER UNA GENERICA FUNZIONE $g(x)$ → NONCHÉ AD UNA TRASFORMAZIONE DELLA R.V STESSA

DISCRETO:

$$E[g(X)] = \sum_{x \in S} g(x) f_X(x)$$

CONTINUO:

$$E[g(X)] = \int_S g(x) \cdot f_X(x)\, dx$$

APPLICARE TRASFORMAZIONI ALLE R.V È UTILE IN QUANTO QUESTE TRASFORMAZIONI PERMETTONO DI ADATTARE I MODELLI PROBABILISTICI ALLA REALTÀ, IN MODI CHE LA VARIABILE ORIGINALE NON PUÒ FARE

ES INVECE DEL LATO VOGLIO ANALIZZARE LA DISTRIBUZIONE DELL'AREA! CONSIDERO X^2...

OPPURE POSSO USARE LA SCALA LOGARITMICA X AGEVOLARMI NEI CALCOLI.

Random Vector - Expectation

QUANDO PASSIAMO DA VARIABILI ALEATORIE UNIDIMENSIONALI A VETTORI ALEATORI, IL CONCETTO DI VALORE ATTESO SI ESTENDE NATURALMENTE. PER UN VETTORE ALEATORIO $X = (X_1, X_2, \dots, X_m)$ IL VALORE ATTESO È IL VETTORE DEI VALORI ATTESI DELLE SINGOLE COMPONENTI

$$E[X] = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_m] \end{pmatrix}$$

È IL CENTRO DI MASSA DELLA FUNZIONE DI PROBABILITÀ NELLO SPAZIO m-DIMENSIONALE

PROPRIETÀ

<u>LINEARITÀ</u>: SE $Y = AX + b$, A È UNA MATRICE e B UN VETTORE COSTANTE

$$E[Y] = E[AX+b] = A \cdot E[X] + b$$

INOLTRE $Y = X + Z$, DOVE X e Z SONO VETTORI ALEATORI:

$$E[Y] = E[X] + E[Z]$$

Covarianza tra R.Variabili

È LA MISURA DI COME 2 VARIABILI ALEATORIE VARIANO INSIEME. È DEFINITA COME:

$$cov(X,Y) = E[XY] - E[X]E[Y] = E\Big[(X - E[X])(Y - E[Y])\Big]$$

SE LE VARIABILI TENDONO A MUOVERSI NELLA STESSA DIREZIONE (QUANDO UNA È SOPRA LA SUA MEDIA, L'ALTRA TENDE AD ESSERE SOPRA LA SUA MEDIA), LA COVARIANZA È POSITIVA SE TENDONO A MUOVERSI IN DIREZIONI OPPOSTE LA COVARIANZA È NEGATIVA SE NON C'È UN PATTERN DI VARIAZIONE CONGIUNTA, LA COVARIANZA È VICINA A 0

IN PARTICOLARE SE $X \perp\!\!\!\perp Y \Rightarrow cov(X,Y) = 0$ (NON È VERO L'OPPOSTO)

$$cov(X,Y) = E_{XY}[XY] - E_X[X]\,E_Y[Y] = cov(X,Y)$$

CHE A VOLTE LA COV È POSITIVA A VOLTE NEGATIVA

SOMMARIAMENTE 0 NON SEGUONO UN TREND

DENTRO LA FORMULA

la covarianza vuole misurare quanto 2 variabili "variano insieme" - cioè quando una variabile si discosta dalla sua media, l'altra tende a discostarsi dalla propria media nella stessa direzione? o in direzione opposta??

$$cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

Valore atteso del prodotto delle deviazioni delle 2 variabili

① Se X è sopra la sua media $(X-E[X]>0)$ e Y lo è anche $(Y-E[Y]>0)$
il prodotto è positivo

② Se $X-E[X]<0$ e pure Y il prodotto è positivo

③ Se X e Y tendono a deviare in direzione opposte è negativo

$$cov(X,Y) = \sum_{x \in S_X} \sum_{y \in S_Y} xy \cdot f_{X,Y}(x,y) - \left(\sum_{x \in S_X} x f_X(x)\right)\left(\sum_{y \in S_Y} y f_Y(y)\right)$$

| oppure

$$\sum_{x \in S_X} \sum_{y \in S_Y} (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)$$

(caso discreto $\int_{\mathbb{R}}$ al posto di Σ)

Correlazione di Pearson

[non si usa x individuare relazioni non lineari tipo x^2]

$$\rho = corr(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)\,var(Y)}} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Dove:
$cov(X,Y)$ è la $cov(X,Y)$
σ_X e σ_Y sono le deviazioni standard

$$corr(X,Y) = \frac{E[(X-E[X])(Y-E[Y])]}{\sqrt{E[(X-E[X])^2]\,E[(Y-E[Y])^2]}}$$

La si può considerare come una covarianza normalizzata. Mentre la covarianza dipende dall'unità di misura (se misuriamo in m invece che cm) la cov cambia. La corr è adimensionale varia sempre -1 e 1 indipendentemente dalle unità di misura.

Se la corr è
- $+1$ perfetta corr positiva quando una ↗ lo fa pure l'altra
- -1 perfetta corr negativa. quando una aumenta l'altra diminuisce proporzionalmente
- 0 non mostrano tendenza a variare insieme in modo lineare

Proprietà

Correlazione perfetta $\text{corr}(X,Y) = 1 \iff Y = aX + b$, dove $a \neq 0$
dove se $a > 0$ la corr è $+1$ se < 0 è -1

Non transitività

Simmetria $\text{corr}(X,Y) = \text{corr}(Y,X)$

Invarianza alla scala: moltiplicare X o Y x una costante positiva non cambia la corr

Invarianza alla traslazione: sommare una una costante a X o Y non cambia la corr.

Matrice di covarianza Σ x r.v.

Estende il concetto di varianza al caso multidimensionale, permettendo di analizzare come le componenti di un vettore aleatorio variano insieme

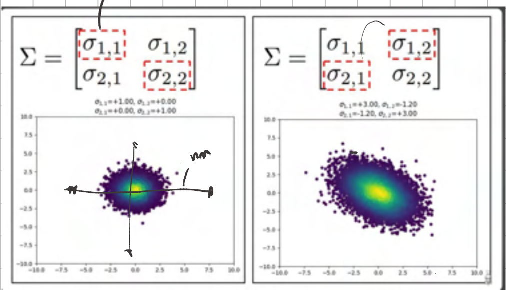
$X = \{X_1, X_2, \dots, X_m\}$ la sua matrice di covarianza è definita come

$$\Sigma = \text{cov}(X) = E[(X-E[X])(X-E[X])^T]$$

$$= \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1,X_2) & \cdots & \text{cov}(X_1,X_m) \\ \text{cov}(X_1,X_2) & \text{var}(X_2) & \cdots & \text{cov}(X_2,X_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_m,X_1) & \text{cov}(X_m,X_2) & \cdots & \text{var}(X_m) \end{bmatrix}$$

varianza lungo le 2 dimensioni

correlazione



↳ se una cambia l'altra come varia??
ci dice come i valori cambiano insieme

le persone alte tendono ad essere anche più pesanti

Proprietà

Simmetria: $\Sigma = \Sigma^T$, perché $cov(X_i,X_j) = cov(X_j,X_i)$

Semidefinita Positiva: Tutti gli autovalori λ_i sono non negativi $(\lambda_i \geq 0)$

Forma Quadratica Non Negativa: , $v^T \Sigma v \geq 0$ $\forall v$ — la forma quadratica da come risultato un numero $\geq$ che rappresenta un ellissoide
(praticamente tutti i vettori che soddisfano l'equazione $v^T \Sigma v$ rappresentano un elissoide)

In statistica se A è una Σ e x un vettore unità la forma quadratica rappresenta la varianza dei dati nella direzione di x, x questo è sempre positivo
le direzioni dove questo valore è grande sono quelle di massima variazione

Trasformazione Lineare della matrice di covarianza

$$\text{if } Y = AX + b \qquad cov(Y) = A\, cov(X) A^T$$

Si sposa perfettamente con quello appreso prima, la moltiplicazione x una matrice A non è altro che una trasformazione dei dati non che un cambio dei vettori di base dello spazio vettoriale $\hat{i}\hat{j}$...

Distribuzioni Discrete di Random Variable

Distribuzione di Bernulli

utilizzata per modellare l'esito di un evento binario, come il lancio di una singola moneta (potenzialmente sbilanciata) che può risultare in testa (H) o croce (T).

Definiamo $X \in [0,1]$ come la nostra variabile randomica, $X=1$ rappresenta il successo e accade con probabilità p
$X=0$ rappresenta il fallimento che ha probabilità $1-p$

A sua volta $X \sim Ber(p)$

Funzione di Massa di Probabilità $\quad P_X(x) = f_X(x) = Ber(x|p) = \begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases} =$

$$= p^x(1-p)^{(1-x)} \Rightarrow \begin{array}{ll} x=0 & P_X(x) = 1-p \\ x=1 & P_X(x) = p \end{array}$$

A volte si scriverà brevemente $P(1) = p$ [sottointendendo $P(X=1)$]

Distribuzione Binomiale

È l'estensione naturale della distribuzione di Bernoulli quando abbiamo più prove indipendenti dello stesso tipo. Modella il numero totale di successi in n esperimenti di Bernoulli indipendenti ciascuno con la stessa probabilità di successo p

ES:

Numero di teste in 10 lanci di moneta

Numero di prodotti difettosi in un lotto

Matematicamente: Consideriamo l'esecuzione di m esperimenti di Bernoulli indipendenti ciascuno con probabilità di successo p, contiamo ora il numero totale di successi X (la R.V rappresenta quindi il numero totale di successi in m Try)

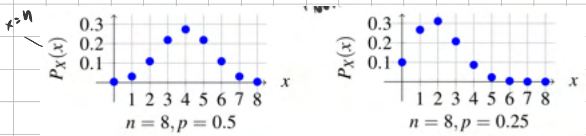
X segue una distribuzione binomiale $X \sim Bin(n,p)$

m è il numero totale di prove

p è la probabilità di successo in ogni singola prova

La funzione di massa di probabilità è $P(X=h) = \binom{m}{h} p^h (1-p)^{m-h}$ dove

$\binom{m}{h} = \frac{m!}{h!(m-h)!}$ è il coefficiente binomiale rappresenta il numero di modi diversi in cui possiamo selezionare h successi tra m prove



Distribuzione Categoriale

Estensione naturale della distribuzione di Bernoulli quando abbiamo più di 2 possibili esiti. Mentre Bernoulli modella eventi binari (come il lancio di una moneta) la categoriale modella il lancio di h possibili risultati (come il lancio di un dado)

Definizione formale

$X \in \{1,2,\dots,h\}$ la R.V può assumere h valori diversi

$X \sim Cat(p)$

$f_X(x) = P(X=x) = p_x = \prod_i p_i^{\mathbb{I}[x=i]}$

Dove:

- $p = (p_1,\dots,p_h)$ è il vettore di probabilità con $\sum_{i=1}^{h} p_i = 1$
- p_i è la probabilità dell'esito i
- $\mathbb{I}$ è la funzione indicatrice $\mathbb{I}[c] = \begin{cases} 1 & \text{if } c \text{ true} \\ 0 & \text{otherwise} \end{cases}$

La funzione indicatrice è come un interruttore che si accende valore 1 quando

UNA CERTA CONDIZIONE È VERIFICATA.

ES: LANCIO UN DADO A 6 FACCE. LA R.V X PUÒ ASSUMERE I VALORI $(1,2,3,4,5,6)$
SE VOGLIO ESPRIMERE LA PROBABILITÀ CHE ESCA 3 $p(X=3)=\frac{1}{6}$

$$p(X=x) = \prod_i p_i^{I[x=i]}$$

PER $i=1$ $\quad I[3=1]=0 \quad p_1^0=1$

$\vdots$

$i=3$ $\quad I[3=3]=1 \quad p_3^1=\frac{1}{6}$

$\vdots$

QUESTA FUNZIONE È DI PER SÉ SCOMODA PER QUESTO SI PREFERISCE USARE UNA CODIFICA ONE-HOT

SE $x=3$, $X=(0,0,1,0,\dots 0)$

QUINDI IN GENERALE

$$x=1 \Rightarrow X=(1,0,\dots 0)$$
$$x=2 \Rightarrow X=(0,1,\dots 0)$$
$$\vdots$$
$$x=n \Rightarrow X=(0,0,\dots 1)$$

IN QUESTO MODO TOLGO L'INDICATORE E USO DIRETTAMENTE

$$f_X(x) = \prod_i p_i^{x_i}$$

DISTRIBUZIONE MULTINOMIALE

LA DISTRIBUZIONE MULTINOMIALE ESTENDE LA DISTRIBUZIONE CATEGORICA ALLO STESSO MODO IN CUI LA BINOMIALE ESTENDE BERNOULLI. IMMAGINA DI ESEGUIRE UN ESPERIMENTO RIPETUTO n VOLTE (LANCIARE UN DADO n VOLTE), DOVE OGNI PROVA PUÒ AVERE k POSSIBILI RISULTATI DIVERSI. ECCO LA DISTRIBUZIONE MULTINOMIALE CI DICE QUALE È LA PROBABILITÀ DI OTTENERE UNA SPECIFICA CONFIGURAZIONE DI RISULTATI

DEFINIZIONE FORMALE

n NUMERO TOTALE DI PROVE INDIPENDENTI
k NUMERO DI POSSIBILI ESITI PER OGNI PROVA
$p=(p_1, p_2, \dots p_k)$ VETTORE DELLE PROBABILITÀ DEI k POSSIBILI ESITI (con $\sum p_i = 1$)
$x=(x_1, x_2, \dots x_k)$ VETTORE DEI CONTEGGI, DOVE x_i RAPPRESENTA QUANTE VOLTE

SI È VERIFICATO L'ESITO i

ALLORA LA PROBABILITÀ DI OTTENERE ESATTAMENTE QUESTA CONFIGURAZIONE DI CONTEGGI È DATA DA:

$$f_X(x) = \frac{n!}{(x_1! \cdot x_2! \ldots \cdot x_n!)} p_1^{x_1} p_2^{x_2} \ldots p_n^{x_n}$$

$$=$$

$$f_X(x) = \frac{n!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} p_i^{x_i}$$

ES LANCIAMO UN DADO A 6 FACCE 10 VOLTE

$n = 10$ (NUMERO DI LANCI)

$k = 6$ (POSSIBILI FACCE DEL DADO)

$p = (1/6, 1/6, \ldots 1/6)$ [PROBABILITÀ UNIFORMI X OGNI FACCIA

VOGLIAMO CALCOLARE LA PROBABILITÀ DI OTTENERE

2 VOLTE 1
1 VOLTA 2
3 VOLTE 3
0 VOLTE 4
2 VOLTE 5
2 VOLTE 6

$\Rightarrow$ QUINDI $X = (2, 1, 3, 0, 2, 2)$ LA CUI SOMMA È 10 OVVERO IL NUMERO TOTALE DI LANCI

$$P(X = (2,1,3,0,2,2)) = \frac{10!}{2!\,1!\,3!\,0!\,2!\,2!} \times (1/6)^2 (1/6) (1/6)^3 (1/6)^0 (1/6)^2 \ldots = 0{,}0025$$

IL COEFFICIENTE MULTINOMIALE: $\frac{n!}{x_1!\, x_2! \ldots x_n!}$ RAPPRESENTA IL NUMERO DI MODI DIVERSI IN CUI POSSIAMO OTTENERE LA CONFIGURAZIONE X SPECIFICATA.

DISTRIBUZIONE GAUSSIANA o NORMALE

IL FONDAMENTO DELLA STATISTICA

UNA VARIABILE ALEATORIA X SEGUE UNA DISTRIBUZIONE NORMALE CON PARAMETRI μ (mu) e σ^2 (SIGMA AL QUADRATO) SE LA SUA DENSITÀ DI PROBABILITÀ È:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- IL TERMINE ESPONENZIALE $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ CREA LA CARATTERISTICA FORMA A CAMPANA, CHE DECRESCE RAPIDAMENTE MAN MANO CHE CI ALLONTANIAMO DA μ IN ENTRAMBE LE DIREZIONI.
- IL FATTORE $\frac{1}{\sqrt{2\pi\sigma^2}}$ È UNA COSTANTE DI NORMALIZZAZIONE CHE GARANTISCE CHE L'AREA

SOTTO LA CURVA SIA 1, COME RICHIESTO X OGNI DENSITÀ DI PROBABILITÀ

① μ (mu): È LA MEDIA DELLA DISTRIBUZIONE, IL CENTRO DELLA CAMPANA, IL VALORE PIÙ PROBABILE NONCHÉ PIÙ DENSO. $E[X] = \mu$

σ (SIGMA): LA DEVIAZIONE STANDARD, LA MISURA DI QUANTO I VALORI SONO "SPARSI" ATTORNO ALLA MEDIA. MAGGIORE È σ + LARGA È LA CAMPANA. LA VARIANZA È $\sigma^2 = Var(X)$

λ (LAMBDA): $1/\sigma^2$, CHIAMATO PRECISIONE, CHE È L'INVERSO DELLA VARIANZA. MAGGIORE È LA PRECISIONE, PIÙ STRETTA È LA DISTRIBUZIONE ATTORNO ALLA MEDIA.

NORMALE STANDARD ($N(0,1)$)

QUANDO UNA VARIABILE SEGUE UNA DISTRIBUZIONE NORMALE $\mu = 0$ e $\sigma^2 = 1$, DICIAMO CHE È UNA NORMALE STANDARD.

- QUALSIASI VARIABILE NORMALE $X \sim N(\mu, \sigma^2)$ PUÒ ESSERE STANDARDIZZATA USANDO LA TRASFORMAZIONE $z = (x-\mu)/\sigma$, OTTENENDO $Z \sim N(0,1)$

LA NORMALIZZAZIONE MI PERMETTE DI CONFRONTARE PIÙ R.V. GAUSSIANE DIVERSE

ES: VOGLIO CONFRONTARE QUANTO SONO ECCEZIONALI

- UN BAMBINO ALTO 140 cm — IN UNA CLASSE $\mu = 130$ cm e $\sigma = 5$ cm
- UN ADULTO CHE PESA 90 kg - IN UNA POPOLAZIONE - $\mu = 75$ kg $\sigma = 12$ kg

X IL BIMBO $z = (140-130)/5 = 2$

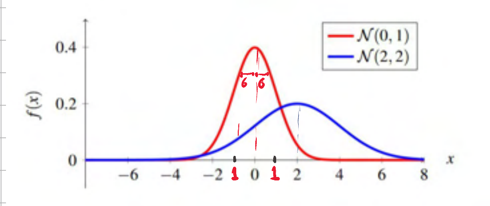
X L'ADULTO $z = (90-75)/12 = 1{,}25$

L'ALTEZZA DEL BAMBINO È PIÙ ECCEZIONALE SI DISCOSTA DI PIÙ DAL VALORE MEDIO = 0

È DISTANTE PIÙ STANDARD DEVIATION

Caratteristiche Grafiche.

La curva è perfettamente simmetrica rispetto alla media



Varianza più grande corrisponde ad un appiattimento maggiore

Teorema del Limite Centrale

Questo teorema spiega perché la distribuzione normale appare così influente in natura e nella scienza

Se sommiamo molte variabili casuali indipendenti (qualunque sia la loro distribuzione originale) la somma tenderà ad avere una distribuzione approssimativamente normale.

/ indipendenti e identicamente distribuite

Se prendiamo una sequenza di random variable i.i.d $X_1, X_2, ..., X_n$ ciascuna con media μ e varianza $\sigma^2 > 0$ formiamo la loro somma $S_n = X_1 + X_2 + ... + X_n$ e formiamo la loro somma $S_n = X_1 + X_2 + ... X_n$ ⇒ la variabile normalizzata data dalla somma di tutte le r.v avrà una distribuzione del tipo (si può partire da qualsiasi distribuzione, basta che sia la stessa)

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

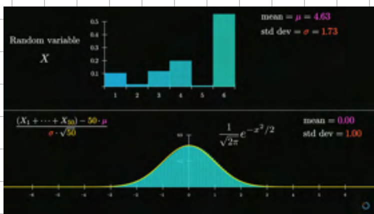
converge ad una normale standard quando $n \to \infty$

Nota

- $n\mu$ si sposta a destra man mano che aggiungiamo r.V. [$n\mu$ è la media della nuova distribuzione $\downarrow S_n$]
- $Var(x+y) = Var(x) + Var(y)$ [dato che X e Y sono indipendenti]

$Var(x) = Var(y) = \sigma^2$

$\sigma = \sqrt{Var(x+y)} = \sqrt{Var(x) + Var(y)} = \sqrt{2\sigma^2} = \sqrt{2}\,\sigma$ nella formula $\sqrt{n}\,\sigma$ è la nuova deviazione standard



MULTIVARIATE GAUSSIAN DISTRIBUTION.

LA DISTRIBUZIONE GAUSSIANA MULTIVARIATA È UNA GENERALIZZAZIONE DELLA DISTRIBUZIONE NORMALE A DIMENSIONI MULTIPLE. LA GAUSSIANA MULTIVARIATA DESCRIVE LA DISTRIBUZIONE CONGIUNTA DI PIÙ VARIABILI ALEATORIE CORRELATE

PARTIAMO DAL CASO PIÙ SEMPLICE UN VETTORE X COMPOSTO DA VARIABILI NORMALI STANDARD, INDIPENDENTI
$X=[X_1, X_2 \dots X_n]^T$ DOVE OGNI $X_i \sim N(0,1)$ E TUTTE LE VARIABILI SONO INDIPENDENTI TRA LORO

IN QUESTO CASO LA FUNZIONE DI PROBABILITÀ CONGIUNTA È SEMPLICEMENTE IL PRODOTTO DELLE DENSITÀ INDIVIDUALI

$$f_X(x) = \prod_{i=1}^{N} f_{X_i}(x_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \cdot e^{(-(x_i)^2/2)} = (2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}x^T x}$$

DICIAMO CHE X SEGUE UNA DISTRIBUZIONE NORMALE MULTIVARIATA STANDARD, OVVERO $X \sim N(0, I)$
0 È IL VETTORE DELLE MEDIE, I È LA MATRICE DI IDENTITÀ NONCHÉ MATRICE DI COVARIANZA

GENERALE

PER INVECE CONSIDERARE IL CASO GENERALE, DI UNA MVG CON MEDIA μ E MATRICE DI COVARIANZA Σ, POTREMO PENSARE DI APPLICARE UNA TRASFORMAZIONE LINEARE AL VETTORE NORMALE STANDARD
$X = AY + \mu$

DOVE
$Y = N(0, I)$ VETTORE NORMALE STANDARD MULTIVARIATO
A È UNA MATRICE TALE CHE $\Sigma = AA^T$
μ VETTORE DELLE MEDIE.

$$X \sim N(\mu, \Sigma)$$

LA FUNZIONE DI DENSITÀ DI PROBABILITÀ DIVENTA:

$$f_X(x) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

$|\Sigma|$ È IL DETERMINANTE DELLA MATRICE DI COVARIANZA
$|\Sigma^{-1}|$ È L'INVERSA ANCHE CHIAMATA Λ MATRICE DI PRECISIONE

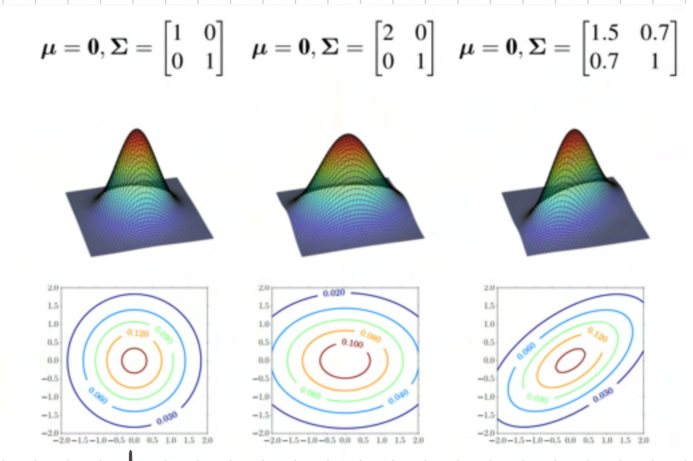
LA FUNZIONE DI DENSITÀ DELLA DISTRIBUZIONE NORMALE MULTIVARIATA HA UNA FORMA A CAMPANA IN n DIMENSIONI.
LE CURVE DI LIVELLO, (PUNTI CON UGUALE DENSITÀ) FORMANO ELLISSOIDI CENTRATI IN μ.

L'ORIENTAMENTO E LA FORMA DI QUESTI ELLISSOIDI, SONO DETERMINATI

DA Σ:

- Gli autovettori di Σ determinano la direzione degli assi principali dell'ellissoide (SVD(PCA))

- Gli autovalori di Σ determinano la lunghezza degli assi principali dell'ellissoide



Gli ellissoidi sono centrati perché le variabili hanno la stessa varianza

Variabili indipendenti ma la prima ha una varianza maggiore (allungamento asse X)

Le variabili hanno covarianza positiva quindi la densità è concentrata nel primo e terzo quadrante.

Gli ellissoidi sono inclinati a causa della relazione positiva ρ (se una è sopra media, lo è pure l'altra.)

Density Estimation

La stima della probabilità permette di inferire le probabilità di eventi futuri basandoli su osservazioni passate.

<u>Lancio di una moneta</u>

Immagino di avere una moneta di cui non conosciamo il grado di equilibrio. Potrebbe essere equa o sbilanciata. Abbiamo però osservato m lanci di questa moneta e vogliamo prevedere la probabilità di ottenere testa nel prossimo lancio

I risultati dei lanci osservati sono $(x_1, x_2, \dots x_n)$ dove $x_i = 1$ [lancio i-esimo ha dato testa]

$x_i = 0$ [lancio i-esimo ha dato croce]

Il nostro obiettivo è stimare $P(X_t = 1 \mid X_1 = x_1, \dots X_n = x_n)$ cioè la probabilità che il prossimo lancio X_t dia testa dato che abbiamo osservato i risultati $x_1 \dots x_n$

Se conoscessimo π ovvero la probabilità di ottenere testa, potremmo modellare ciascun lancio come una variabile aleatoria di Bernoulli con parametro π

$X_i \sim Ber(\pi)$

Poiché in questo caso i lanci sono i.i.d. anche il prossimo lancio seguirebbe la stessa distribuzione

$X_t \sim Ber(\pi)$

Quindi $P(X_t = 1) = \pi$ [indipendentemente dai risultati precedenti]

Noi però non conosciamo π. L'approccio frequentista assume che esiste questo valore "un valore vero" π_t che determina la probabilità di testa, e che dobbiamo stimare questo valore dai dati.

↓

Per fare ciò si usa la massima verosimiglianza (Maximum Likelihood): cerchiamo il valore di π che rende più probabile l'osservazione dei dati raccolti

La **likelihood** è la probabilità di osservare i dati $x_1, \dots x_n$ condizionata dal valore π (sapendo che ogni singolo lancio ha prob π)

$$L(\pi) = P(X_1 = x_1, \dots X_n = x_n \mid \pi)$$

I lanci sono indipendenti quindi $L(\pi) =$ al prodotto delle probabilità dei singoli lanci $= \prod_{i=1}^{n} p(X_i = x_i \mid \pi)$

x la distribuzione di Bernoulli (considerando $p = \pi$) che è la nostra incognita/condizione $= \prod_{i=1}^{n} \pi^{x_i} (1-\pi)^{1-x_i}$

$$p(X_i = x_i \mid \pi) = \pi^{x_i} (1-\pi)^{1-x_i}$$

Dobbiamo ora trovare il valore π che massimizza questa funzione

[Nota la likehood è di per sé una probabilità - [rappresenta in modo per spiegare quei dati] - quanto π è compatibile con i miei dati [più è compatibile più la likehood è alta]]

Per trovare il π che ci massimizza la funzione è conveniente lavorare con i log il (log-likelihood)

$$\ell(\pi) = \log L(\pi) = \sum_{i=1}^{N} \left[x_i \log \pi + (1-x_i) \log(1-\pi) \right]$$

Il logaritmo è una funzione monotona crescente quindi il massimo è lo stesso di $\ell(\pi)$

Calcoliamo la derivata $\frac{d\ell}{d\pi} = \sum_{i=1}^{N} \frac{x_i}{\pi} - \frac{(1-x_i)}{1-\pi} = 0 \Rightarrow \hat{\pi}_{ml} = \frac{1}{m} \sum_{i=1}^{m} x_i$

Molto intuitivo. La stima di massima verosimiglianza x π è la media dei valori x_i (la proporzione di teste osservate)

La probabilità del prossimo lancio è quindi $P(X_t = 1 | X_1 = x_1 \dots X_m = x_m) = \hat{\pi}_{ml}$

Limitazioni

Se abbiamo osservato solo 3 lanci, la nostra stima sarebbe $\hat{\pi}_{ml} = 3/3 = 1$ [non uscirà mai sempre testa, i dati sono limitati]

L'approccio frequentista ignora l'incertezza associata alla stima $\hat{\pi}_{ml}$

con pochi dati la nostra stima potrebbe essere + lontana dal vero valore π

Stima bayesiana della densità di probabilità [non nel dettaglio]

Nell'approccio bayesiano trattiamo il parametro π non come un valore fisso e sconosciuto ma come una variabile aleatoria (r.v) con una propria distribuzione di probabilità.

Prima di osservare qualsiasi dato, abbiamo una distribuzione a priori $f_\pi(\pi)$ che rappresenta la nostra conoscenza o credenza iniziale su π, x esempio se pensiamo che la moneta sia equa potremo scegliere una distribuzione iniziale centrata su π

Se pensiamo al lancio della moneta, esprimiamo la nostra incertezza sul valore di π attraverso una distribuzione di probabilità.

Prima di lanciare la moneta abbiamo già delle idee su quanto sia probabile che π assuma certi valori. Questa è la nostra "distribuzione a priori" $f_\pi(\pi)$

Se pensiamo che la moneta sia equa potremo usare una distribuzione con un picco intorno a 0.5 ....

Lanciamo la moneta n volte e otteniamo h teste e $t = n-h$ croci

Calcoliamo la verosimiglianza ovvero, la probabilità di ottenere esattamente i dati osservati dato un certo valore di π

$$L(\pi) = P(dati/\pi) = \pi^{h}(1-\pi)^{t}$$

Il teorema di Bayes ci permette adesso di "invertire" la probabilità: non vogliamo $P(dati/\pi)$ ma $P(\pi/dati)$, cioè data l'evidenza dei nostri lanci, quanto è probabile che π abbia un certo valore?

$$p(\pi/dati) = \frac{P(dati/\pi)\, P(\pi)}{P(dati)}$$

Più in generale, se consideriamo l'osservazione dei dati $x_1, \dots, x_n$, aggiorniamo la nostra conoscenza su π utilizzando il teorema di Bayes, considerando π come una R.V con distribuzione a priori $f_\pi(\pi)$

La nota densità di probabilità nel caso discreto $P(X=x)$

$$f_{\pi|X_1\dots X_m}(\overset{\text{verso}}{\pi / x_1 \dots x_m}) = \frac{\overset{\text{inverso}}{P(X_1=x_1, \dots X_m=x_m|\pi)}\, f_\pi(\pi)}{P(X_1=x_1 \dots X_m=x_m)} = ①$$

Ricordati

Posso vedere $P(X_1=x_1 \dots X_m=x_m)$ come una probabilità marginale della funzione di probabilità congiunta $f_{XY}(x,y) \to f_{X\pi}(x,\pi) \Rightarrow P(X_1=x_1, \dots X_m=x_m, \Pi=\pi)$

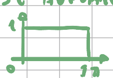
Inoltre la probabilità condizionata è pari a $f_{X|Y}(x/y) = \frac{f_{XY}(x,y)}{f_Y(y)}$

$$\Rightarrow f_{X,Y}(x,y) = f_{X|Y}(x/y)\, f_Y(y)$$

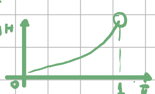
$$① \frac{P(X_1=x_1, \dots X_m=x_m|\pi)\, f_\pi(\pi)}{\int P(X_1=x_1, \dots X_m=x_m, \Pi=\pi)\, d\pi} = \frac{\overset{g(\pi)}{P(X_1=x_1, \dots X_m=x_m|\pi)}\, f_\pi(\pi)}{\int P(X_1=x_1 \dots X_m=x_m/\pi)\, f_\pi(\pi)\, d\pi}$$

Il denominatore agisce anche come una costante di normalizzazione (la densità di probabilità deve essere valida)

Nota: se avevamo supposto che



dopo aver osservato i dati



1 è la soluzione più probabile ma ho altri valori comunque possibili per π

Calcolare la distribuzione a posteriori usandola x fare previsioni sul prossimo lancio X_t

$xt = 1, 0, \ldots$

$$P(Xt = xt | X_1 = x_1 \ldots X_m = x_m) = \int P(Xt = xt | \pi) f_{\pi | X_1 = x_1 \ldots X_m = x_m}(\pi | x_1 \ldots x_m) d\pi$$

QUESTO INTEGRALE PUÒ ESSERE INTERPRETATO COME UNA MEDIA PONDERATA DELLE PREVISIONI, CONDIZIONATE A OGNI POSSIBILE VALORE DI π, DOVE I PESI SONO DATI DALLA DISTRIBUZIONE A POSTERIORI

LA PREVISIONE TIENE CONTO DELL'INTERA DISTRIBUZIONE DI π, INCORPORANDO L'INCERTEZZA SULLA STIMA DEL PARAMETRO

LIMITAZIONI E VANTAGGI

- INCORPORA CONOSCENZE A PRIORI
- GESTISCE L'INCERTEZZA → FUNZIONA BENE CON DATI LIMITATI

MA

- LA SCELTA DELLA DISTRIBUZIONE A PRIORI È SOGGETTIVA
- COMPLESSITÀ COMPUTAZIONALE

STIMA DELLA DENSITÀ PER VARIABILI CONTINUE: IL CASO GAUSSIANO

LA DISTRIBUZIONE GAUSSIANA (O NORMALE) EMERGE NATURALMENTE IN MOLTI CONTESTI PER DIVERSE RAGIONI

① TEOREMA DEL LIMITE CENTRALE

② OSSERVAZIONE EMPIRICA: MOLTI FENOMENI NATURALI PRESENTANO ISTOGRAMMI CON FORME SIMILI ALLA CURVA A CAMPANA, CARATTERISTICA DELLA GAUSSIANA. ES ALTEZZA DI UNA POPOLAZIONE ...

③ TRATTABILE MATEMATICAMENTE: LA DISTRIBUZIONE GAUSSIANA HA PROPRIETÀ MATEMATICHE CHE LA RENDONO RELATIVAMENTE FACILE DA MANIPOLARE NELL'ANALISI STATISTICA.

È RAGIONEVOLE QUINDI MODELLARE I NOSTRI DATI COME CAMPIONI DI UNA DISTRIBUZIONE GAUSSIANA.

SUPPONIAMO DI AVERE UN INSIEME DI DATI $D = (x_1, x_2, \ldots x_m)$ E DECIDIAMO DI MODELLARE QUESTI DATI COME CAMPIONI DI UNA DISTRIBUZIONE GAUSSIANA CON MEDIA μ E VARIANZA V O PRECISIONE $\lambda = \frac{1}{V}$.

ASSUMIAMO CHE I PUNTI SIANO STATI GENERATI DA VARIABILI ALEATORIE INDIPENDENTI E IDENTICAMENTE DISTRIBUITE $X_1, X_2, \ldots X_m$. DATA UNA SPECIFICA

SCELTA DEI PARAMETRI $\theta = (\mu, v)$; LA DENSITÀ DI PROBABILITÀ DI CIASCUNA X_i È

SE CONOSCIAMO I PARAMETRI

$$f_{X|\theta}(x) = N(x|\mu, v) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-\mu)^2}{2v}}$$

NEL CASO DISCRETO È QUINDI POSSIBILE DEFINIRE LA FUNZIONE DI VEROSIMIGLIANZA PER I PARAMETRI θ COME IL PRODOTTO DELLE DENSITÀ INDIVIDUALI.

$$L(\theta) = \prod_{i=1}^{n} f_{X_i|\theta}(x_i) = \prod_{i=1}^{n} N(x_i|\mu, v)$$

ESPRIME LA PROBABILITÀ DI OTTENERE I DATI $x_1 \ldots x_m$, CONDIZIONATI DAL VALORE θ

CONSIDERANDO L'APPROCCIO FREQUENTISTA SI ASSUME CHE ESISTANO DEI VALORI "VERI" MA SCONOSCIUTI PER I PARAMETRI DEL MODELLO, DENOTATI COME μ^T E v^T. SE CONOSCESSIMO QUESTI VALORI, LA DENSITÀ PER I NUOVI CAMPIONI X_t SAREBBE SEMPLICEMENTE:

$$f_{X_t}(x_t) = N(x_t|\mu_T, v_T)$$

MA DATO CHE NON CONOSCIAMO QUESTI VALORI, BISOGNA STIMARLI A PARTIRE DAI DATI DISPONIBILI

STIMATORI STATICI

UNO STIMATORE È UNA FUNZIONE T CHE MAPPA IL DATASET D A VALORI STIMATI PER I PARAMETRI DEL MODELLO θ^*:

$$\theta^* = T(D)$$

NEL CASO GAUSSIANO SI VUOLE STIMARE SIA LA MEDIA μ CHE LA VARIANZA v

INOLTRE VOGLIAMO CHE GLI STIMATORI SIANO CONSISTENTI, CHE CONVERGANO (IN PROBABILITÀ) AI VERI VALORI DEI PARAMETRI μ^T E v^T MAN MANO CHE LA DIMENSIONE DEL CAMPIONE $m \to \infty$

IN GENERALE SI USANO 2 METODI PRINCIPALI IL (MOM) METODO DEI MOMENTI E LA MAXIMUM LIKELIHOOD (ML)

METODO DEI MOMENTI (MOM)

SI FANNO CORRISPONDERE I MOMENTI TEORICI DELLA DISTRIBUZIONE (COME MEDIA E VARIANZA) AI MOMENTI EMPIRICI CALCOLATI A PARTIRE DAI DATI.

PER UNA DISTRIBUZIONE GAUSSIANA I PRIMI 2 MOMENTI SONO μ E LA VARIANZA v. GLI STIMATORI MOM SONO

$$\mu^*_{mom} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$v^*_{mom} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu^*_{mom})^2$$

C.V.D

- LA STIMA DELLA MEDIA È SEMPLICEMENTE LA MEDIA ARITMETICA DEI DATI
- LA STIMA DELLA VARIANZA È LA MEDIA DEI QUADRATI DELLE DEVIAZIONI DELLA MEDIA STIMATA.

SONO STIMATORI CONSISTENTI CONVERGONO AI VALORI REALI QUANDO LA DIMENSIONE DEL CAMPIONE CRESCE ALL'INFINITO.

STIMA DI MASSIMA VEROSOMIGLIANZA (ML)

LA STIMA DI MASSIMA VEROSOMIGLIANZA CERCA I VALORI DEI PARAMETRI CHE MASSIMIZZANO LA PROBABILITÀ DI OSSERVARE I DATI CHE ABBIAMO EFFETTIVAMENTE OSSERVATO. TROVA I PARAMETRI CHE RENDONO I NOSTRI DATI PIÙ PROBABILI.

LA DERIVAZIONE DEGLI STIMATORI ML È PIÙ COMPLESSA.

1. SCRIVERE LA FUNZIONE LIKELIHOOD PER UN CAMPIONE DI n OSSERVAZIONI iid.

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

DOVE $\theta = (\mu, v)$ SONO I PARAMETRI E f_X È LA DENSITÀ GAUSSIANA.

2. PER SEMPLIFICARE I CALCOLI, LAVORIAMO CON IL LOGARITMO DELLA VEROSOMIGLIANZA

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f_X(x_i|\theta)$$

IL LOGARITMO DELLA DENSITÀ PER LA GAUSSIANA HA UNA FORMA SEMPLICE

$$\log N(x|\mu, v) = \log \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-\mu)^2}{2v}} = \log(2\pi v)^{-\frac{1}{2}} + \log e^{-\frac{(x-\mu)^2}{2v}}$$

$$= -\frac{1}{2}\log(2\pi v) - \frac{(x-\mu)^2}{2v}\underbrace{\log e}_{=1} = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log v - \frac{(x-\mu)^2}{2v}$$

$\gamma = \frac{1}{\lambda}$ $\quad \varepsilon_1 + \frac{1}{2}\log\lambda - \frac{\lambda}{2}(x-\mu)^2$

ε_1 È UNA COSTANTE CHE NON DIPENDE DAI PARAMETRI.

$$\ell(\theta) = \sum_{i=1}^{n} \log N(x|\mu,\lambda^{-1}) = \varepsilon_2 + \sum_{i=1}^{n}\left[\frac{1}{2}\log\lambda - \frac{\lambda}{2}(x_i-\mu)^2\right] =$$

$$= \varepsilon_2 + \frac{n}{2}\log\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

ESPANDENDO I TERMINI QUADRATICI

$$= \varepsilon_2 + \frac{n}{2}\log\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}(x_i^2+\mu^2-2x_i\mu) = \varepsilon_2 + \frac{n}{2}\log\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}x_i^2$$

$$+ \frac{n\lambda\mu^2}{2} - 2\mu\frac{\lambda}{2}\sum_{i=1}^{n}x_i$$

$$\ell(\theta) = \varepsilon_2 + \frac{n}{2}\log\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}x_i^2 + \lambda\mu\sum_{i=1}^{n}x_i - \frac{n\lambda\mu^2}{2}$$

PER TROVARE I VALORI DI λ E μ CHE MASSIMIZZANO LA VEROSIMIGLIANZA CALCOLIAMO LE DERIVATE PARZIALI RISPETTO A μ E λ E LE PONIAMO = A 0

$$\frac{\partial\ell}{\partial\mu} = \lambda\sum_{i=1}^{n}x_i - 2n\frac{\lambda}{2}\mu = 0 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n}x_i - n\mu = 0 \Rightarrow \mu = \frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\frac{\partial\ell}{\partial\lambda} = \frac{n}{2\lambda} - \frac{1}{2}\sum_{i=1}^{n}x_i^2 + \mu\sum_{i=1}^{n}x_i - \frac{n\mu^2}{2} = 0$$

$$\Rightarrow \quad \frac{n}{\lambda} = \sum_{i=1}^{n}x_i^2 - 2\sum_{i=1}^{n}x_i\mu + \sum_{i=1}^{n}\mu^2$$

$$\lambda = \frac{n}{\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)}$$

$$v = \frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

NEL CASO DELLA DISTRIBUZIONE GAUSSIANA GLI STIMATORI MAP E ML QUINDI COINCIDONO. [QUESTO RISULTATO NON VALE X TUTTE LE DISTRIBUZIONI]

CONCLUDENDO IL NUOVO VALORE CHE CERCHIAMO x_t È INDIVIDUATO DA

$$f_{X_t|X_1 \dots X_n}(x_t | x_1 \dots x_n) \approx N(x_t | \mu_{ML}, v_{ML})$$

IL PROSSIMO VALORE x_t AVRÀ UN VALORE DEFINITO DALLA SUA DISTRIBUZIONE GAUSSIANA CON UNA CERTA PROBABILITÀ.