

PROBABILITÀ E STATISTICA

Riassunto Appunti Personali
dalle Lezioni Uninettuno

a.a. 2014 - 2015

Il presente documento è da intendersi “AS IS”.
Salvo errori e errori di stampa.

PROBABILITÀ CON IL MODO

DI FIDUCIA : SONO DISPOSTO A PAGARE
UN p VICINO A 1 SE HO FIDUCIA ; SE NON HO
FIDUCIA PAGO UN p VICINO A 0

pago p per un evento E
 \uparrow
S.p., con S. valore di x che si realizza

CON QUESTI
SEMI IO SONO
LO SCAMBIARE
INVERTITI DA
VIL RAC

$E=1 = \text{vero} \Rightarrow \text{GUADAGNO } G(E) = 1 - p$

$E=0 = \text{falso} \Rightarrow \text{GUADAGNO } G(E^c) = -p$

e una perdita

PER LA COERENZA DEVE VALERE :

$G(E) \cdot G(E^c) \leq 0$ (i due guadagni non
devono avere lo
stesso segno)

da cui

$$0 \leq p \leq 1$$

Per una scommessa coerente lo somma
da pagare per entrare nella scommessa è
sempre un numero compreso tra 0 e 1.
Questo ci porta alla definizione naturale
di probabilità come grado di fiducia. [Lotto
3 e 4]

PROBABILITÀ DELL'EVENTO E
COME MISURA DEL GRADO DI

FIDUCIA IN E ; LA PROBABILITÀ SARÀ
ESPRRESSA MEDIANTE UN NUMERO REALE

$p = P(E)$; p sarà preso in modo tale

che una ipotetico scommessa su E con p

(e importo S risultano, che non influente p)
sia coerente:

$$0 \leq P(E) \leq 1$$

$$P(E) = p$$

$$P(\emptyset) = 0$$

$$P(\Omega) = 1$$

PER STABILIRE IL VALORE DI p
OCORRE PARLARE DI EVENTI e di
EVENTI CONDIZIONATI

EVENTO: E INDIVIDUATO DA
UNA PROPOSIZIONE

• IMPLICAZIONE DI EVENTI

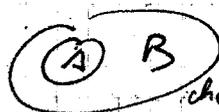
$$A \subseteq B$$

↳ convenuto in

SE $A \in \text{VERO}$ ANCHE

$B \in \text{VERO}$

SE $A \subseteq B$ e $B \subseteq A$ allora $B = A$



ogni volta che A è vero, B è vero.

EVENTO CERTO = SEMPRE VERO = Ω , $|\Omega| = 1$

EVENTO IMPOSSIBILE = SEMPRE FALSO = \emptyset , $|\emptyset| = 0$

EVENTO CONTRARIO = E^c VERO SE $E = \text{FALSO}$
 FALSO SE $E = \text{VERO}$

• UNIONE DI EVENTI

$$A \cup B$$

OR

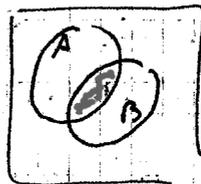


almeno uno dei due eventi si verifica

• INTERSEZIONE DI EVENTI

$$A \cap B$$

AND



A e B si verificano contemporaneamente.

PROPRIETA':
 COMMUTATIVA, ASSOCIATIVA
 DISTRIBUTIVA

$$\left. \begin{aligned} A \cup B &= B \cup A \\ A \cap B &= B \cap A \end{aligned} \right\} \text{COMMUTATIVA}$$

$$\begin{aligned} A \vee (B \vee C) &= (A \vee B) \vee C \\ A \wedge (B \wedge C) &= (A \wedge B) \wedge C \end{aligned} \left. \vphantom{\begin{aligned} A \vee (B \vee C) \\ A \wedge (B \wedge C) \end{aligned}} \right\} \text{ASSOCIATIVA}$$

$$\begin{aligned} A \wedge (B \vee C) &= (A \wedge B) \vee (A \wedge C) \\ A \vee (B \wedge C) &= (A \vee B) \wedge (A \vee C) \end{aligned} \left. \vphantom{\begin{aligned} A \wedge (B \vee C) \\ A \vee (B \wedge C) \end{aligned}} \right\} \text{DISTRIBUZIONE}$$

ALTRE PROPRIETÀ:

$$\Omega^c = \emptyset \quad e \quad \emptyset^c = \Omega$$

$$A \vee A = A \quad e \quad A \wedge A = A$$

$$A \vee A^c = \Omega$$

$$A \wedge A^c = \emptyset$$

$$A \vee \Omega = \Omega$$

$$e \quad A \vee \emptyset = A$$

$$\Omega \vee \emptyset = \Omega$$

$$e \quad \Omega \wedge \emptyset = \emptyset$$

$$A \vee (A \wedge B) = A$$

$$A \wedge (A \vee B) = A$$

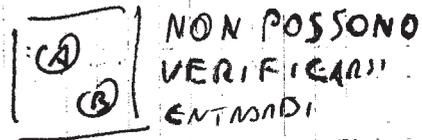
$$(A \vee B)^c = A^c \wedge B^c$$

$$(A \wedge B)^c = A^c \vee B^c$$

} Leggi di De Morgan

EVENTI INCOMPATIBILI

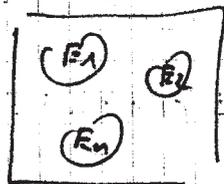
$$A \cap B = \emptyset$$



EVENTI A DUE A DUE INCOMPATIBILI

$$E_1, E_2, \dots, E_n$$

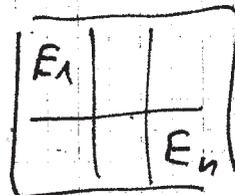
$$E_r \cap E_s = \emptyset, r \neq s$$



PARTIZIONE DI Ω

E_1, E_2, \dots, E_n a due a due incompatibili

$$\bigcup_{i=1}^n E_i = \Omega$$



(L'unione degli eventi è Ω)

(sono due a due incompatibili, ma ricoprono tutto Ω)

• EVENTI INDIPENDENTI

(IL VERIFICARSI DI UNO NON INFLUENZA L'ALTRO)
(ESTRAZIONE CON RESTITUZIONE)

• A e B indipendenti, allora

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A|B) = P(A) \quad (\text{esempio, palline bianche e nere})$$

$$P(B|A) = P(B) \quad (\text{esempio, 1° e 2° estrazione con restituzione})$$

• EVENTI DIPENDENTI

(estrazione senza restituzione)

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

(esempio, palline bianche e nere)

• EVENTI INCOMPATIBILI

(sempre eventi elementari)

$$A \cap B = \emptyset, \quad P(A \cup B) = P(A) + P(B)$$

• EVENTI COMPATIBILI

$$A \cap B \neq \emptyset, \quad P(A \cup B) = P(A) + P(B) + P(A \cap B)$$

PROBABILITA' - ASSIOMI

(i) $0 \leq P(E) \leq 1$

(ii) $P(\Omega) = 1$, $P(\emptyset) = 0$

(iii) (E_1, E_2, \dots, E_n) ^{partizione di eventi} _{mutuamente esclusivi} ^{in cui uno e solo evento si verifica}

$$\sum_{i=1}^n P(E_i) = P\left(\bigvee_{i=1}^n E_i\right) = \underbrace{1}_{P(\Omega)}$$

probabilità dell'unione

TEOREMA DELLE PROBABILITA' TOTALI
SE GLI EVENTI SI DICHIARANO A DUE A DUE INCOMPATIBILI E NON PARTIZIONE, IL RISULTATO $\neq 1$ PERCHÉ DEFINIRI NON PARLANO NON È TUTTO Ω [scorrettamente, p. 18]

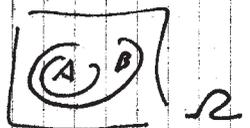
PROBABILITA' TOTALI

E_1, E_2, \dots, E_n ^{due a due incompatibili} _{e non partizione}

$$\sum_{i=1}^n P(E_i) = P\left(\bigvee_{i=1}^n E_i\right) \neq 1 \text{ , perché non è tutto } \Omega$$

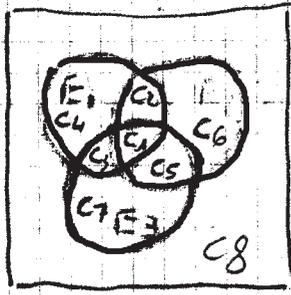
MONOTONIA

$$A \subseteq B \Rightarrow P(A) \leq P(B)$$



I COSTITUENTI

SONO NUOVI EVENTI OTTENUTI CON LE INTERSEZIONI DEGLI EVENTI E DEI LORO CONTRARI.



Ω

Gli eventi

$C_1 \dots C_8$ si

chiamano costituenti.

$$C_1 = E_1 \cap E_2 \cap E_3$$

$$C_2 = E_1 \cap E_2 \cap E_3^c$$

$$C_3 = E_1 \cap E_2^c \cap E_3$$

$$C_4 = E_1 \cap E_2^c \cap E_3^c$$

$$C_5 = E_1^c \cap E_2 \cap E_3$$

$$C_6 = E_1^c \cap E_2 \cap E_3^c$$

$$C_7 = E_1^c \cap E_2^c \cap E_3$$

$$C_8 = E_1^c \cap E_2^c \cap E_3^c$$

C_1 è vero quando sono veri E_1, E_2, E_3

Tutti i costituenti sono possibili due a due
 almeno possibile e la loro unione è certo

* Se $P(E_1) = a$, valore conosciuto a dato

allora $d_1 + d_2 + d_3 + d_4 = a$

↑
 valore di C_1 → se la probabilità

Se $P(E_2) = b$, allora

$$d_1 + d_2 + d_3 + d_4 = b$$

che sia vero E_1 e non E_2 $d_2 + d_3 + d_4$

Se $P(E_3) = c$, allora

$$d_1 + d_3 + d_5 + d_7 = c$$

Dove vale, per le coordinate

$$d_1 + d_2 + d_3 + d_4 + d_5 + d_6 + d_7 + d_8 = 1$$

Per risolvere il sistema

$$\left. \begin{array}{l} d_k \geq 0 \text{ per} \\ \text{tutti} \\ \text{gli } k \end{array} \right\} \text{ per verificare la}$$

$$\left\{ \begin{array}{l} d_1 + d_2 + d_5 + d_6 = d \\ d_1 + d_2 + d_3 + d_4 = b \\ d_1 + d_3 + d_5 + d_7 = c \\ d_1 + d_2 + \dots + d_8 = 1 \end{array} \right.$$

si determinano i valori di d che sono possibili.

Se gli d_k sono tutti ≥ 0 e la somma è 1 allora l'assegnazione è coerente, viceversa, non lo è.

* Perché $E_2 = C_1 \vee C_2 \vee C_5 \vee C_6$

NUMERO ALFATORIO DISCRETO

$$X = x_1 I_{E_1} + x_2 I_{E_2} + \dots + x_n I_{E_n}$$

X è detto numero aleatorio discreto, perché può assumere un numero finito di valori possibili

E_1, E_2, \dots, E_n sono n eventi qualunque

x_1, x_2, \dots, x_n sono n numeri reali qualunque

$I_{E_1}, I_{E_2}, \dots, I_{E_n}$ sono gli indicatori degli eventi che possono assumere valore 0 o 1.

Il numero aleatorio X è dato da una sommatoria, la somma dei coefficienti agli eventi veri:

$$X = x_{i_1} + x_{i_2} + \dots + x_{i_r}$$

X , numero aleatorio, sarà la somma di alcuni, eventualmente tutti, i coefficienti x_i , quelli per cui E_i è vero.

L'indicatore di un evento è un

particolare numero aleatorio;
per $n=2$ abbiamo infatti

$$X = x_1 |E_1| + x_2 |E^c| \quad \text{con } E_1 = E \\ \text{e } E_2 = E^c$$

Un numero aleatorio può essere moltiplicato
per un numero reale; due numeri aleatori
possono essere sommati.

In generale l'insieme dei numeri
aleatori forma uno spazio vettoriale.

Se gli eventi sono una partizione
allora i numeri aleatori si dicono
in forma canonica.

NUMERI ALEATORI IN FORMA CANONICA

Gli eventi sono a due a due incompatibili e
la loro unione è l'evento certo.

$$E_i \wedge E_k = \emptyset \quad (i \neq k)$$

$$E_1 \vee E_2 \vee \dots \vee E_n = \Omega \quad (\text{partizione})$$

$$X = x_1 |E_1| + x_2 |E_2| + \dots + x_n |E_n|$$

Uno ed uno solo degli eventi è vero.

Generalizzazione del concetto di numero aleatorio discreto, i cui valori possono essere un numero infinito ma numerabile, in forma canonica.

partizione di Ω : $\{E_1, E_2, \dots, E_n\}$
finita o numerabile

successione a : $\{a_1, a_2, \dots, a_n\}$
di numeri reali.

Si definisce numero aleatorio discreto X :

$$X = \sum_{k=1}^n a_k |E_k|$$

in cui un solo addendo sopravvive allo somma.
Cioè un solo evento è vero ma non sappiamo quale.

Il numero aleatorio generalizza il concetto di evento in quanto l'indicatore di un evento è un particolare numero aleatorio.

LA PREVISIONE

Dato un numero aleatorio generale X

$$X = x_1 |E_1| + x_2 |E_2| + \dots + x_n |E_n|$$

si definisce previsione di un numero aleatorio X (o valore atteso, o valor medio, o speranza matematica):

$$IP(X) = p_1 x_1 + p_2 x_2 + \dots + p_n x_n$$

$$\text{con } p_i = P(E_i), \quad i = 1, 2, \dots, n$$

Il concetto di previsione generalizza il concetto di probabilità.

Caso particolare $X = |E|$ probabilità di E

$$IP(X) = IP(|E|) = 1 \cdot P(E) + 0 \cdot (1 - P(E)) = P(E)$$

e la previsione si riduce alla probabilità.

Altre proprietà:

$$IP(2X) = 2 IP(X)$$

$$IP(X+Y) = IP(X) + IP(Y)$$

$$IP(X_1 + \dots + X_n) = IP(X_1) + \dots + IP(X_n)$$

$$IP(X) = \frac{x_1 + \dots + x_n}{n} \quad \text{media aritmetica}$$

Il numero di successi:

$$Y = |\Lambda_1| + \dots + |\Lambda_n|$$

$$IP(Y) = p_1 + \dots + p_n \quad \text{previsione dei successi}$$

Λ_i equiprobabili

$$IP(Y) = np$$

Frequenza dei successi

$$V = \frac{Y}{n}$$

$$IP(V) = \frac{1}{n} IP(Y) = \frac{p_1 + \dots + p_n}{n}$$

Λ_i equiprobabili

$$IP(V) = p_i \quad \text{probabilità}$$

previsione della
frequenza in caso
di eventi
equiprobabili.

VARIANZA

- Sia $\mu = IP(X)$ la previsione del numero elettronico X , cioè

$IP(X) = x_1 |E_1| + x_2 |E_2| + \dots + x_n |E_n|$ dove E_i sono una partizione

-
$$var(X) = \underbrace{p_1}_{\text{probab. } E_1} (x_1 - \mu)^2 + p_2 (x_2 - \mu)^2 + \dots + p_n (x_n - \mu)^2$$

Sia $\sigma = \sqrt{var(X)}$ lo scarto quadratico medio (o scarto standard).

Possiamo definire

-
$$Y = (X - \mu)^2 = (x_1 - \mu)^2 |E_1| + \dots + (x_n - \mu)^2 |E_n|$$

$$var(X) = IP(Y) = IP((X - \mu)^2)$$

Lo scarto quadratico medio
La varianza di X è una previsione, non di X , ma di Y .

- La varianza ha lo scopo di affiancare alle previsioni un altro indice che riporti come sono dispersi i valori di X

rispetto alla previsione.

La varianza dà una misura della dispersione.

Se è piccola i valori di X , cioè (x_1, x_2, \dots, x_n) sono vicini al suo valore medio.

Varianza grande \Rightarrow maggiore incertezza
(maggiore dispersione)

Varianza piccola \Rightarrow minore incertezza

La varianza di un indicatore

di un evento E è pq , cioè

$$\text{var}(|E|) = pq, \text{ massima a } p = \frac{1}{2}$$

PROPRIETÀ DI PREVISIONE E VARIANZA

$$a = X = a | \Omega | \quad IP(a) = a$$

$$\text{var}(X) = IP(X^2) - (IP(X))^2$$

$$IP(c + X) = c + IP(X)$$

$$\text{var}(c + X) = \text{var}(X)$$

$$\text{var}(x+x) = \text{var}(x)$$

IL NUMERO ALEATORIO STANDARDIZZATO

$$Y = \frac{X - \mu}{\sigma}$$

$$IP(Y) = 0$$

$$\text{var}(Y) = 1$$

LA COVARIANZA

$$\text{cov}(X, Y) = IP(XY) - IP(X)IP(Y)$$

$$\text{cov}(X, X) = \text{var}(X)$$

$$\text{cov}(X, Y) = IP[(X - IP(X))(Y - IP(Y))]$$

(pressione del prodotto
degli scarti)

La covarianza può assumere valori
positivi, negativi e nulli. Non è additivo

$$\text{var}(x+y) = \text{var}(x) + \text{var}(y) + 2\text{cov}(x, y)$$

$$\begin{aligned} \text{var}(x_1 + \dots + x_n) &= \\ &= \text{var}(x_1) + \dots + \text{var}(x_m) + 2 \sum_{\{i_1, i_2\}} \text{cov}(x_{i_1}, x_{i_2}) \end{aligned}$$

IL COEFFICIENTE DI CORRELAZIONE
DI X e Y (una covarianza normalizzata)

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (\text{prodotto degli scarti standard})$$

$$\sigma_X = \sqrt{\text{var}(X)}, \quad \sigma_Y = \sqrt{\text{var}(Y)}$$

$$-1 \leq \rho(X, Y) \leq 1$$

Si ha $\rho^2(X, Y) = 1$ se e solo se
fra X e Y c'è un legame funzionale
lineare del tipo $Y = bX + d$ con $b \neq 0$

PROBABILITÀ CONDIZIONATA

$$p = P(E|H)$$

E dato H

$$E|H \begin{cases} \text{Vero} & (E \cap H \text{ vero}) \\ \text{Falso} & (E^c \cap H \text{ vero}) \\ ? & (H \text{ falso}) \end{cases} \quad \left. \vphantom{E|H} \right\} H \text{ sempre vero}$$

$$- p \begin{cases} 1 & (E|H \text{ vero}) \\ 0 & (E|H \text{ falso}) \\ p & (E|H ?) \end{cases}$$

L'evento condizionato è quando $H = \Omega$

L'evento condizionato è un particolare numero elastico e 3 valori:

$$E|H = X = 1 \cdot |E \cap H| + 0 \cdot |E^c \cap H| + p \cdot |H^c|$$

I valori di X possono essere:

$$1, 0, p$$

↓
coincide con
la somma che
prevediamo

$$p = IP(X) = P(E \cap H) + p - p P(H)$$

da cui $P(E \cap H) = p P(H)$

TEOREMA DELLE PROBABILITÀ COMPORTE

$$P(E \cap H) = P(E|H) P(H) = P(H|E) P(E)$$

$$P(E|I) = p$$

probabilità
condizionata

Se $P(E) > 0$ allora

$$P(I|E) = \frac{P(I \cap E)}{P(E)}$$

che è la rappresentazione diretta
della probabilità condizionata, in
funzione di due probabilità
non condizionate.

Abbiamo le seguenti formule:

$$P(E) = \sum_{k=1}^n P(I_k) P(E|I_k)$$

della formula di disintegrazione
di $P(E)$ rispetto alla partizione $\{I_k\}$

Il concetto di probabilità condizionata
consente di valutare probabilità in
corrispondenza a diversi stati di informazione

TEOREMA DI BAYES verosimiglianza

$$P(H_r | E) = P(H_r) \cdot \frac{P(E | H_r) \text{ elementi di una partit.}}{\sum_{r=1}^n P(H_r) P(E | H_r)}$$

In base a questa relazione possiamo aggiornare le probabilità.

$P(H_r)$ è la probabilità iniziale, che viene aggiornata una volta entrati in possesso della informazione E .

Teorema delle probabilità composte:

$$P(H_r) P(E | H_r) = P(E) P(H_r | E)$$

e, la formula di disintegrazione

$$P(E) > 0$$

$$P(H_r | E) = P(H_r) \frac{P(E | H_r)}{P(E)}$$

$$E = E \wedge \Omega = E \wedge \left(\bigvee_{k=1}^n H_k \right) = \bigvee_{k=1}^n (H_k \wedge E)$$

INDIPENDENZA STOCASTICA ^{ACCIDENTARIA} DEGLI EVENTI

ESSA DIPENDE DAI DUE EVENTI E DALLE LORO PROBABILITA'

$P(E)P(H) > 0$, DUE EVENTI E e H HANNO
PROBABILITA' POSITIVE

E e H SONO STOCASTICAMENTE INDIPENDENTI
SE

$$P(E \cap H) = P(E)P(H)$$

INOLTRE, SE E e H SONO INDIPENDENTI,

$$P(H|E) = P(H) \text{ e } P(E|H) = P(E)$$

CIO'E' SE H E' INDIPENDENTE DA E ALLORA E
E' INDIPENDENTE DA H .

DUNQUE L'INFORMAZIONE E NON MODIFICA LA
PROBABILITA' DI H E VICEVERSA

SE DUE EVENTI SONO STOCASTICAMENTE
INDIPENDENTI, LO SONO ANCHE TUTTI
QUELLI OTTENUTI METTENDO IN TUTTI I
NON POSSIBILI IL CONTRARIO.

SE $P(E_1 \cap E_2) = P(E_1)P(E_2)$ ALLORA SONO
INDIPENDENTI ANCHE E_1^c, E_2 ; E_1, E_2^c ; E_1^c, E_2^c .

CIOE' TUTTE LE COPPIE OTTENUTE
METTENDO IL CONTRARIO AD UNO O
ENTRABI GLI EVENTI SONO ANCORA
STOCSTICAMENTE INDIPENDENTI.

FAMIGLIE DI EVENTI

INDIPENDENTI

Famiglia, finita o numerabile, di eventi E
con $0 < P(E) < 1$. Gli eventi della famiglia
sono stocsticamente indipendenti se, qualunque
sia k e comunque si scelgano k eventi
 E_1, E_2, \dots, E_k della famiglia, deve valere:

$$P(E_1 \wedge E_2 \wedge \dots \wedge E_k) = P(E_1) \cdot P(E_2) \cdot \dots \cdot P(E_k)$$

PROBLEMI DELLE PROVE RIPETUTE

PAG 63 ESCORPANO

Dati n eventi indipendenti ed equiprobabili.

$$E_1, E_2, \dots, E_n \quad P(E_i) = p \quad i = 1, 2, \dots, n$$

quale è la probabilità che n ne verifichino
 h , con $0 \leq h \leq n$?

$E = h$ successi su n prove con $0 \leq h \leq n$

$$P(E) = \binom{n}{h} p^h q^{n-h} \quad \text{con } q = 1 - p$$

Se $E = n$ primi h successi, allora $P(E) = p^h q^{n-h}$. LEZ 9

ESTRAZIONI DA URNE pag. 63 ESCOZZESE

ESTRAZIONI CON RESTITUZIONE

E = "h successi su n prove" ad es. h palline

E_j = "pallina bianca alla j-esima estrazione" $j=1, 2, \dots, n$
bianche su n estrazioni

$$P(E_j) = \frac{pN}{N} = p$$

$$E = \bigvee_{j_1, \dots, j_h} A_{j_1, \dots, j_h}^{(n)} \quad \text{con} \quad A_{j_1, \dots, j_h}^{(n)} = E_{j_1} \wedge \dots \wedge E_{j_h}$$

cioè E è unione di $\binom{n}{h}$ eventi incompatibili $\wedge E_{j_1} \dots$

$$P(E) = \binom{n}{h} p^h q^{n-h} \quad \text{con} \quad q = 1-p$$

E = "h successi su n prove", ad es. h palline bianche su n estrazioni con restituzione.
Assumo indipendenza ed equiprobabilità.

ESTRAZIONI SENZA RESTITUZIONE

Gli eventi E_j ($j=1, 2, \dots, n$) sono equiprobabili ma non sono indipendenti:

$$P(E_j) = \frac{pN}{N} = p \quad \text{equiprobabili}$$

$$P(E_1 \wedge E_2) = p \frac{pN-1}{N-1} \neq P(E_1) P(E_2) = p^2$$

non indipendenti

L'ipotesi di indipendenza fra $P(E_1 \cap E_2)$ e $P(E_1) \cdot P(E_2)$ sussiste solo se $p=1$ che è il caso banale in cui l'urna contiene tutte palline bianche.

$E =$ "h successi su n prove"

ed r le palline bianche in n estrazioni senza restituzione

$$P(E) = \frac{\binom{pN}{h} \binom{qN}{n-h}}{\binom{N}{n}}$$

stessa formula con

$r = pN$, urna con N palline, r bianche e rimanenti nere.

La probabilità che lo n -uplo estratto contenga h palline bianche è:

$$\frac{\binom{r}{h} \binom{N-r}{n-h}}{\binom{N}{n}}$$

Si osserva che la probabilità di verificarsi l'evento fissato (e di non verificarsi dei rimanenti $n-h$) dipende solo da quanti e non da quali eventi si considerano.

ESTIMAZIONE DA URNA

DI COMPOSIZIONE IGNOTA

ϑ_r = frazione incognita di palline bianche

Numero $r = \vartheta_r N$ con $r = 0, 1, 2, \dots, N$ = numero palline bianche.

N = numero totale palline

H_r = evento "le palline bianche nell'urna sono $r = \vartheta_r N$ "

Possiamo calcolare la probabilità E_j condizionata a H_r , che sarà quella della composizione che ha r palline bianche, mediante la formula di disintegrazione.

$$P(E_j | H_r) = \vartheta_r = \frac{r}{N} \quad (j = 1, 2, \dots, N)$$

$$E_j = E_j \wedge \Omega = E_j \wedge \left(\bigvee_{r=0}^N H_r \right) = \bigvee_{r=0}^N (E_j \wedge H_r),$$

$$\begin{aligned} P(E_j) &= \sum_{r=0}^N P(E_j \wedge H_r) = \sum_{r=0}^N P(H_r) P(E_j | H_r) \\ &= \sum_{r=0}^N P(H_r) \cdot \frac{r}{N} \end{aligned}$$

Quindi

$E =$ "h successi in n prove"

dal es. h palline bianche in n estrazioni da urna di composizione ignota con restituzione

$$\begin{aligned} P(E) &= \sum_{r=0}^N P(H_r) P(E|H_r) = \\ &= \sum_{r=0}^N P(H_r) \binom{n}{h} \left(\frac{r}{N}\right)^h \left(1 - \frac{r}{N}\right)^{n-h} = \\ &= \sum_{r=0}^N P(H_r) \binom{n}{h} \varphi_r^h (1 - \varphi_r)^{n-h} \end{aligned}$$

ESTRAZIONE SENZA RESTITUZIONE
DA URNA DI COMPOSIZIONE IGNOTA

PROCEDURA ANALOGA si riconosce la stessa formula
PER ESTRAZIONE SENZA RESTITUZIONE, con φ_r al posto di p :

$$P(E|H_r) = \frac{\binom{\varphi_r N}{h} \binom{(1-\varphi_r)N}{n-h}}{\binom{N}{n}}$$

Usando le formule di disintegrazione otteniamo:

$$P(E) = \sum_{r=0}^N P(H_r) \frac{\binom{\varphi_r N}{h} \binom{(1-\varphi_r)N}{n-h}}{\binom{N}{n}} \quad \text{LEB. 10}$$

DISTRIBUZIONE BINOMIALE

La distribuzione di probabilità è:

Trovare la probabilità che il numero aleatorio X sia uguale al valore x .

Trovare cioè che il numero dei successi sia uguale ad un certo valore dato.

$P(X=x)$ probabilità di valori possibili

Quando la probabilità viene trovata per tutti i valori possibili di x , cioè in tutto il codominio allora è stata costruita la distribuzione di probabilità del numero aleatorio

Siano A_k ($k=1, 2, \dots, n$) eventi indipendenti ed equiprobabili, con probabilità p .

Sia X il numero dei successi:

$$X = |A_1| + |A_2| + \dots + |A_n|$$

Con queste ipotesi, eventi indipendenti, equiprobabili, con probabilità p , la distribuzione che nasce, cioè la probabilità che X sia uguale a x , si chiama distribuzione binomiale.

Il valore della distribuzione binomiale è:

$$P_{Xn} = P(X=x) = \binom{n}{x} p^x q^{n-x}$$

x successi su n prove

$$0 \leq x \leq n$$

Il calcolo della previsione è:

$$IP(X) = np$$

condizionale della
previsione applicata

La distribuzione binomiale si applica ad es. alle estrazioni con restituzione da urna di composizione nota.

DISTRIBUZIONE

IPERGEOMETRICA

Siano N eventi A_k ($k=1,2,\dots,N$) e si conosca il numero r di quelli veri ($0 \leq r \leq N$).

La probabilità di verificarsi di h eventi fissati fra n di essi (e di non verificarsi dei rimanenti $n-h$) dipende solo da come siano h e n con $h \leq n \leq N$ e non dai particolari eventi A_1, \dots, A_N considerati.

In sostanza, se ho 15 eventi, so' che 7 sono veri e ne prendo 12 ($n=12$), quale e' la probabilita' che fra i 12, 5 siano veri?

In queste condizioni gli eventi sono equiprobabili.

$$P(A_k) = p = \frac{r}{N}$$

Se Y e' il numero esatto dei successi, la sua distribuzione di probabilita' (cioe' che $Y=y$) si chiama distribuzione ipergeometrica.

Il suo valore e' :

$$2_{xn}^{(nr)} = P(Y=x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

La previsione di Y e' :

$$|P(Y) = np = \frac{nr}{N}$$

Si osservi che la distribuzione binomiale e la distribuzione ipergeometrica sono completamente diverse, ma hanno entrambe la stessa previsione...

L'urna di composizione nota è un modello per queste distribuzioni:

A_k = la k -esima pallina estratta è bianca

ESTRAZIONE CON RESTITUZIONE: GLI EVENTI A_k SONO INDIPENDENTI ED EQUIPROBABILI E SI APPLICA LA DISTRIBUZIONE BINOMIALE

ESTRAZIONE SENZA RESTITUZIONE: GLI EVENTI A_k SONO EQUIPROBABILI MA NON SONO INDIPENDENTI E ALLORA È APPLICABILE IL MODELLO DELLA DISTRIBUZIONE IPERGEOMETRICA.

La distribuzione binomiale è approssimabile con la ipergeometrica con il limite che tende all'infinito del calcolo dei successi in n prove fissate in n prove nel caso senza restituzioni.

Il valore che tende all'infinito è il numero di casi possibili N .

Mantenendo fissi n , numero di eventi che considero come sottoinsieme del numero totale N , e mantenendo fisso il rapporto

$\frac{r}{N} = p$, con r numero di casi favorevoli, la binomiale è uguale al seguente limite:

$$\lim_{N \rightarrow \infty} P(A_{j_1, j_2, \dots, j_h}^{(n)}) = p^h q^{n-h}$$

Si osserva che $\frac{r}{N} = p$, e $N \rightarrow \infty$ anche $r \rightarrow \infty$.

Assieme dunque quelle che si chiamano l'approssimazione binomiale della ipergeometrica. Infinitamente possiamo dire che se il numero di palline nell'urna, la composizione nota, è enormemente grande rispetto al numero di estrazioni che facciamo allora non c'è molta differenza fra estrazioni con o senza restituzione.

VARIANZA nella distribuzione IPERGEOMETRICA

$$Y = |A_1| + |A_2| + \dots + |A_n|$$

$$\text{var}(Y) = \sum_{k=1}^n \text{var}(|A_k|) + 2 \sum \text{cov}(|A_{i_1}|, |A_{i_2}|)$$

il valore è:

$$\text{var}(|A_k|) = pq$$

Tutte le possibili
combinazioni
a due a due

$$\text{cov}(|A_{i_1}|, |A_{i_2}|) = p \frac{r-1}{N-1} - p^2$$

la covarianza è la previsione del prodotto meno il prodotto della previsione.

Abbiamo dunque che :

$$\begin{aligned} \text{var}(Y) &= npq + 2 \cdot \frac{n(n-1)}{2} \left(p \frac{r-1}{N-1} - p^2 \right) = \\ &= npq \left(1 - \frac{n-1}{N-1} \right) \end{aligned}$$

VARIANZA della distribuzione BINOMIALE

$$\text{var}(X) = npq \quad \left(= \sum_{k=1}^n \text{var}(I_{A_k}) \right)$$

In pratica :

IPERGEOMETRICA

BINOMIALE

$$\text{var}(Y) = npq \left(1 - \frac{n-1}{N-1} \right)$$

$$\text{var}(X) = npq \quad (\text{con restituzione})$$

Si nota che

$$\text{var}(Y) \leq \text{var}(X) \quad \text{a causa del fattore "1 - ...".}$$

Ricordando che una variante grande esprime grande incertezza e varianze piccole esprime poca incertezza allora l'incertezza è minore nel caso di estrazione senza restituzione.

Questo perché via via che si estraggono palline dell'urna si diminuisce e capre il risultato.

Ricapitolando : $\text{var}(Y) \leq \text{var}(X)$
 $\text{var}(Y) = \text{var}(X) \quad \text{per } n=1$

LEZ. 11

$\text{var}(Y)$ decresce al crescere di n

DISTRIBUZIONI DISCRETE

IL CASO IN CUI NON È PIÙ FINITO, MA INFINITO;
IN PARTICOLARE INFINITO NUMERABILE.

PROPRIETÀ GENERALI

$$X = \sum_{k=1}^{\infty} x_k \mathbb{1}_{E_k} \quad \text{con } E_k \text{ partizione di } \Omega$$

$$a_k = P(X = x_k)$$

La distribuzione di probabilità

$$IP(X) = \sum_{k=1}^{\infty} a_k x_k \quad \text{se assolutamente convergente (proprietà commutativa) e}$$
$$\sum_{k=1}^{\infty} a_k = 1 \quad \text{la serie delle probabilità deve essere uguale a 1.}$$

DISTRIBUZIONE DI POISSON

DIPENDE DAL PARAMETRO λ

$\lambda > 0$, x interi non negativi

$$a_x = P(X_\lambda = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \forall x \in \mathbb{N}$$

Si applica ad un fenomeno ripetibile al più un certo numero n di volte in un intervallo di tempo T (incidente, nascita, arrivo di un cliente ad uno sportello, chiamata ad un centralino telefonico ecc.)

Il valore della distribuzione si ottiene per approssimazione: vd. i calcoli a pag 76 del testo e a pag. 12.2 degli appunti.

$$IP(X_\lambda) = \sum_{x=0}^{\infty} x a_x = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

serie esponenziale che converge a e^λ

$$\text{var}(X_\lambda) = \lambda$$

DISTRIBUZIONE GEOMETRICA

Si ripensa al fenomeno di successione di eventi A_k ($k=1, 2, \dots, n$) indipendenti ed equiprobabili, con probabilità p , come il lancio illimitato di una moneta.

Il numero casuale che associamo a questa successione è $X =$ numero di prove fino al primo successo

$$P(X=x) = q^{x-1} p \quad x \in \mathbb{N}$$

$$\begin{aligned} IP(X) &= \sum_{x=1}^{\infty} x q^{x-1} p = p \sum_{x=1}^{\infty} x (q)^{x-1} = \\ &= \frac{1}{p} \end{aligned}$$

Una proprietà della distribuzione geometrica è la "proprietà senza memoria".

Per ogni $n, k \in \mathbb{N}$:

$$\begin{aligned} P(X > n+k \mid X > k) &= \frac{P(X > n+k)}{P(X > k)} = \\ &= \frac{q^{n+k}}{q^k} = P(X > n) \end{aligned}$$

La probabilità è indipendente da k , cioè non ha memoria del passato.

Quello che conta è quante prove aspettiamo, par. ed n .

È vero il viceversa: avendo un numero esattorio del quale non sappiamo la distribuzione e sia $P(X=n) = p_n$, con p_n da determinare, vale che

$$P(X > k+n) = P(X > k) P(X > n), \quad \text{cui}$$

ha distribuzione geometrica, vogliamo vedere come è fatto X .

Poniamo

$$Q_{n+1} = P_1 + P_2 + \dots + P_n + P_{n+1}$$

Per $K=1$ si ha $1 - Q_{n+1} = (1 - P_1)(1 - Q_n)$,

$$1 - Q_{n+1} = (1 - P_1)^{n+1} \quad \text{ottenuto iterativamente}$$

ma

$$P_{n+1} = Q_{n+1} - Q_n$$

$$P_{n+1} = \underbrace{(1 - P_1)^n}_{4} P_1$$

e questa è la formula della distribuzione geometrica.

PROBABILITA' NULLE

Se abbiamo una partizione dell'evento certo (Ω) con potenza del continuo, cioè i cui elementi sono tanti quanti gli elementi dei numeri reali, allora gli eventi di probabilità nulla sono "la maggioranza" (eventualmente tutti), indipendentemente da ogni ipotesi di equiprobabilità.

L'assegnazione $P(E_r) = 0$ per ogni evento della partizione, $E_r \in A$, non corrisponde necessariamente ad un giudizio di equiprobabilità.

Viene posta la σ -additività per riuscire a definire la probabilità ovunque.

Ogni sottoinsieme di Ω può rappresentarsi un evento.

NUMERI ALEATORI CONTINUI

- Prendiamo come famiglia di eventi A l'insieme dei numeri reali e vediamo cosa è una distribuzione di probabilità sull'insieme dei reali, cioè la distribuzione continua di probabilità sull'insieme dei reali.
- Sia $A = \mathbb{R}$ una famiglia di eventi. A è rappresentata dai punti x di \mathbb{R} .

DISTRIBUZIONE CONTINUA DI PROBABILITÀ SU \mathbb{R}

- (i) ogni $x \in \mathbb{R}$ rappresenta un evento E_x , con $\bigcup_{x \in \mathbb{R}} E_x = \Omega$ e si ha $P(E_x) = 0$
- (ii) esiste una funzione reale $f \geq 0$ definita su tutto \mathbb{R} tale che, dato A un intervallo $[a, b]$, per cui A è l'evento unione degli eventi rappresentati dai punti dell'intervallo $[a, b]$
 $A = [a, b] = \bigcup_{x \in [a, b]} E_x$ allora la
anchi $(-\infty, +\infty)$ $\int_a^b f(x) dx$
45

probabilità di A , $P(A)$ su Ω può
rappresentare come

$$P(A) = \int_A f(x) dx$$

La funzione f è detta densità di
probabilità.

Essa non solo deve essere positiva o nulla,
ma il suo integrale su tutto \mathbb{R} deve essere
uguale a 1 e questo perché Ω si è
"trasferito" in \mathbb{R} .

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad \left(\begin{array}{l} \text{distribuzione} \\ \text{di probabilità} \end{array} \right)$$

DISTRIBUZIONE CONTINUA

Questa distribuzione di probabilità introduce
un nuovo tipo di numero aleatorio,
quelli che si chiamano numeri aleatori
continui che hanno come dominio C tale che
 $\text{card } C = \text{card } \mathbb{R}$
e dominio della potenza del continuo.

Distribuzione continua di probabilità su \mathbb{R}

(i) ogni $x \in \mathbb{R}$ rappresenta un evento E_x ,
con $\bigcup_{x \in \mathbb{R}} E_x = \Omega$ e $P(E_x) = 0$

(ii) esiste una funzione reale $f \geq 0$ definita
su tutto \mathbb{R} tale che se,

$$A = [a, b] = \bigcup_{x \in [a, b]} E_x$$

$$P(A) = \int_a^b f(x) dx$$

PREVISIONE

$$m = IP(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

valore
possibile probabilità

Per i discreti è una sommatoria

Possiamo, in generale, anche scrivere

$$IP(\varphi(X)) = \int_{-\infty}^{+\infty} \varphi(x) f(x) dx$$

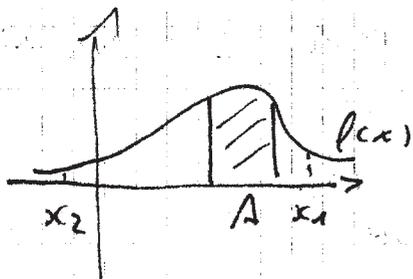
VARIANZA

$$\text{var}(X) = \int_{-\infty}^{+\infty} (x-m)^2 f(x) dx$$

Per le distribuzioni continue la condizione $P(X=x) = 0$ per ogni x non deriva da giudizi di equiprobabilità.

Il valore della densità esprime il grado

di fiducia. La densità deve essere scelta uniforme, cioè costante, se vogliamo dare un giudizio di equiprobabilità.

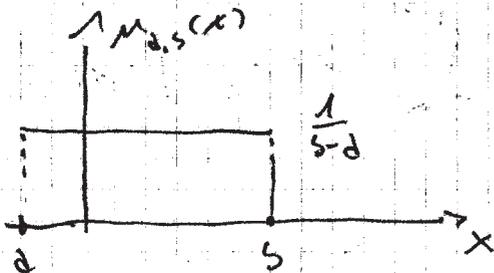


Numero elettorale continuo con distribuzione uniforme sull'intervallo limitato $A = [a, b] \subset \mathbb{R}$; densità di probabilità

$$f_{a,b}(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$$

e l'area del rettangolo deve essere 1.

Si deve avere un intervallo limitato proprio per avere valore 1 dell'area.



DISTRIBUZIONI CONTINUE

QUALUNQUE NUMERO ALEATORIO, DISCRETO O CONTINUO O DI ALTRA TIPO È SEMPRE UNA FUNZIONE DA \mathbb{R} ; L'INSIEME DEGLI EVENTI, IN \mathbb{R} , L'INSIEME DEI REALI

PER ASSEGNARE LA DISTRIBUZIONE CONTINUA BISSA ASSEGNARE LA DENSITA' DI PROBABILITA'.

ATTRAVERSO L'INTEGRALE DELLA DENSITA' ESTENDO A

OPPORTUNI SOTTOINSIEMI DI \mathbb{R} TROVAMO LE

PROBABILITA' DI QUALUNQUE SOTTOINSIEME DI \mathbb{R} CHE CI RAPPRESENTA SEMPRE UN EVENTO.

LA DENSITA' PRENDE IL POSTO DELLA PROBABILITA' NEL CASO DI FIDUCIA

DISTRIBUZIONE BETA

NUMERO ALEATORIO CON DISTRIBUZIONE BETA

Un numero aleatorio si dice con distribuzione beta quando, dati tre parametri positivi reali, $r > 0$, $s > 0$, $x \in (0, 1)$,

l'espressione analitica è:

$$B_{r,s}(x) = \frac{\int_0^x (r+t) \text{ funzione gamma} \cdot x^{r-1} \cdot (1-x)^{s-1} \text{ costante che normalizza a 1}}{\int_0^1 (r+t) \text{ funzione gamma} \cdot x^{r-1} \cdot (1-x)^{s-1} \text{ costante che normalizza a 1}}$$

con $B_{r,s}(x) = 0$ per $x \in (0, 1) \cup 9$

Se $r = \lambda = 1$ la funzione è costante e vale uno quindi la distribuzione beta restituisce la distribuzione uniforme sull'intervallo $(0, 1)$; vale zero fuori dell'intervallo.

La costante normalizza l'espressione a 1 ed è così definita: per ogni $c \in \mathbb{R}^+$,

$$\Gamma(c) = \int_0^{+\infty} x^{c-1} e^{-x} dx, \text{ vale la proprietà}$$

$$\Gamma(c+1) = c \Gamma(c), \text{ che è una formula iterativa,}$$

quindi

$$\begin{aligned} \Gamma(c+1) &= \int_0^{+\infty} x^c e^{-x} dx = x^c (-e^{-x}) \Big|_0^{+\infty} + \\ &+ \int_0^{+\infty} e^{-x} c \cdot x^{c-1} dx = \\ &= c \int_0^{+\infty} x^{c-1} e^{-x} dx = c \Gamma(c) \end{aligned}$$

Se $c \in \mathbb{R}^+$ la formula non è facilmente risolvibile. Lo è se $c \in \mathbb{N}$, così se $c = n$ allora $\Gamma(n+1) = n! \Gamma(1) = n!$

Dunque se $r = \lambda = 1$ si ha una distribuzione uniforme.

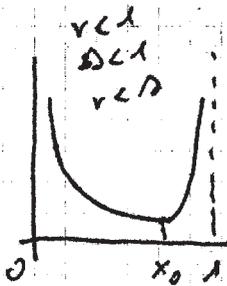
Un modello reale di applicazione può essere la percentuale di persone che hanno una certa caratteristica in una

certa popolazione.

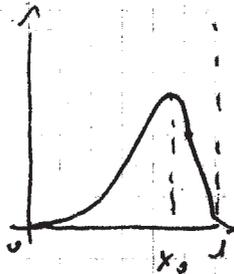
Un qualunque x compreso fra 0 e 1 è la percentuale di pezzi difettosi, in un numero enorme di pezzi N , a sua preferenza, la probabilità di estrarre un pezzo difettoso dal lotto.

Quando $r = N - 1$ densità costante e quindi stesso grado di fiducia. Cioè siamo lo stesso grado di fiducia alla presenza di nessuno, 1, 2, 3, ..., tutti difettosi.

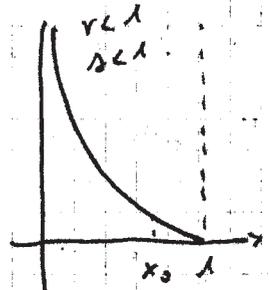
Abbiamo varie costiche di r e s :



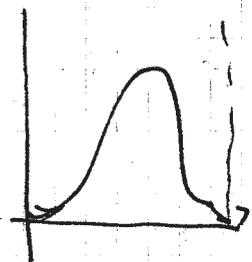
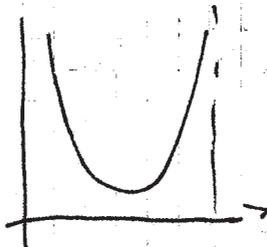
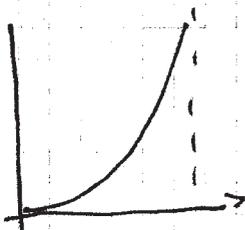
VENGONO PRIVILEGIATI
I DUE CASI ESTREMI,
TUTTI BUONI O TUTTI
DIFETTOSI.



IL NUMERO DI PEZZI
DIFETTOSI È INTORNO
AL VALORE x_0 .



ALTO GRADO DI
FIDUCIA.
SI CREA CI SONO
POCHI PEZZI DIFETTOSI.



Punto di massimo $x_0 = \frac{r+1}{r+\lambda-2}$, $f'(x_0) = 0$; x_0
di minimo in r, λ ca.

Nei altri ha posizioni di minimo.

$$IP(X) = \frac{r}{r+\lambda}$$

$$IP(X^2) = \frac{r(r+1)}{(r+\lambda)(r+\lambda+1)}$$

$$Var(X) = \frac{r\lambda}{(r+\lambda)^2(r+\lambda+1)}$$

DISTRIBUZIONE ESPONENZIALE

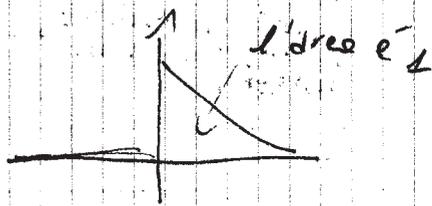
Parametro $\lambda > 0$.

$$f_\lambda(x) = \lambda e^{-\lambda x} \quad \text{per } x \geq 0 \quad e$$

$$f_\lambda(x) = 0 \quad \text{per } x < 0$$

$$IP(X) = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$



DISTRIBUZIONE GAMMA

Parametri: $\lambda > 0$ $c > 0$

$$G_{c,\lambda}(x) = \frac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x} \quad (x > 0);$$

$$G_{c,\lambda}(x) = 0 \quad (x \leq 0)$$

h $c=1$ abbiamo la distribuzione esponenziale.

Al variare di c e λ si ottengono molte forme.

$$\int_{-\infty}^{+\infty} G_{c,\lambda}(x) dx = 1$$

$$IP(x) = \frac{c}{\lambda} \quad \text{var}(x) = \frac{c}{\lambda^2}$$

FUNZIONE DI RIPARTIZIONE

È UNO STRUMENTO CHE PERMETTE DI GESTIRE LE DISTRIBUZIONI DI PROBABILITÀ.

PER LE DISTRIBUZIONI DISCRETE LO STRUMENTO È LA PROBABILITÀ.

PER LE DISTRIBUZIONI CONTINUE LO STRUMENTO È LA DENSITÀ, CON L'INTEGRALE SULLI INSIEMI DEI QUALI INTERESSA LA PROBABILITÀ.

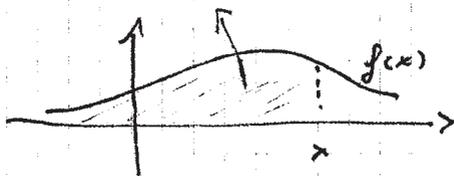
LO STRUMENTO UNIFICANTE CHE VALE INDIPENDENTEMENTE DAL TIPO DI DISTRIBUZIONE (CONTINUA, DISCRETA ED ALTRI TIPI) È LA FUNZIONE DI RIPARTIZIONE, CHE SI RIFERISCE AD UN NUMERO ALEATORIO X ARBITRARIO.

ESSA È DEFINITA SU TUTTO \mathbb{R} .

DATO UN NUMERO ALEATORIO X ARBITRARIO

$$F(x) = P(X \leq x) \quad x \in \mathbb{R}$$

$$\int_{-\infty}^x p(y) dy = P(X \leq x)$$



$F'(x) = f(x)$, la derivata della funzione di ripartizione è la densità

La $f(x)$, la densità, fuori del dominio è 0, per

distribuzioni continue.

Nel caso discreto, come il dado, abbiamo alcuni valori possibili del numero aleatorio X : per il dado tali valori sono 1, 2, 3, 4, 5, 6.

Funzione di ripartizione
numero aleatorio X arbitrario

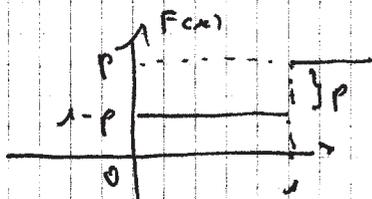
$$F(x) = P(X \leq x) \quad x \in \mathbb{R}$$

x continuo: F derivabile allora
 $F'(x) = f(x)$ con f densità di
probabilità di X .

Se X è discreto (σ-additività numerabile,
inutile nel caso finito, come il dado):

F funzione a scalini
discontinua nei valori possibili x_n di X
costante fra due valori successivi
"salti" uguali a $\Delta_n = P(X = x_n)$

La funzione di ripartizione relativa è un
evento \bar{e} :



In generale abbiamo :

$$x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$$

perché $(X \leq x_1)$ implica $(X \leq x_2)$

Posto $(a, b] = \{X \in (a, b]\}$ abbiamo

$$P((a, b]) = F(b) - F(a)$$

per calcolare la probabilità di un intervallo.

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

$$\lim_{x \rightarrow +\infty} F(x) = 1$$

LA DISTRIBUZIONE

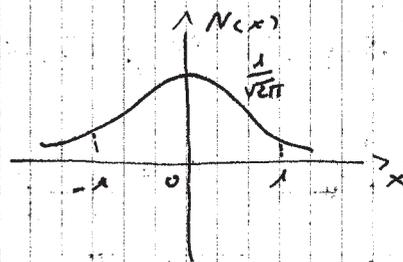
NORMALE o GAUSSIANA

È UNA DISTRIBUZIONE CONTINUA QUINDI PER ASSEGNARLA

DOBBIAMO VERIFICARE CHE È PATA LA DENSITÀ.

QTA PER CUI COMMO TO TUTTO \mathbb{R} ED È DEFINITA SEMPRE POSITIVA.

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



$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$$

le f. non ha primitive

$$IP(X) = 0$$

$$var(X) = 1$$

QUESTA È LA
DISTRIBUZIONE

numero aleatorio standardizzato,
"di zero".

NORMALE STANDARD

La distribuzione normale generica
dipende da due parametri, m e σ
reale, con $\sigma > 0$ e ha infinità due.

Il grafico è uguale a quello dello standard
ma è traslato di m rispetto all'origine.

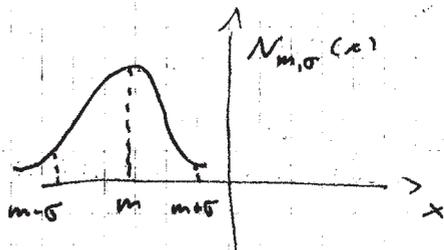
Il punto massimo sta in m e ci sono due punti di flesso calcolabili come media degli estremi, $m - \sigma$ e $m + \sigma$, che nel caso standard erano -1 e 1 .
 La funzione si scosta dal valor medio di una quantità pari allo scarto standard.

Un numero aleatorio $X_{m,\sigma}$ si dice con distribuzione normale quando la sua densità di probabilità è la funzione:

$$N_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \quad x \in \mathbb{R}$$

m , valor medio

σ , scarto standard



La distribuzione normale generale è legata a quella standard (in cui $m=0$ e $\sigma=1$) dalla seguente relazione:

$$N_{m,\sigma}(x) = \frac{1}{\sigma} N\left(\frac{x-m}{\sigma}\right)$$

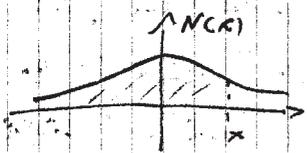
$$IP(X_{m,\sigma}) = m$$

$$var(X_{m,\sigma}) = \sigma^2$$

La funzione di ripartizione Φ della distribuzione continua normale è:

$$\Phi(x) = \int_{-\infty}^x \underbrace{N(\gamma)}_{\text{densità}} d\gamma$$

L'integrale non è calcolabile in quanto N non ha primitiva e si usano delle tavole numeriche della funzione $\Phi(x)$. Per ogni valore di x troviamo il valore della funzione di ripartizione che rappresenta un'area



$\Phi(x)$ vuol dire la probabilità alla sinistra del punto x . Tutta l'area è 1.

La funzione di ripartizione $\Phi_{m,\sigma}$ della normale non standard è:

$$\Phi_{m,\sigma}(x) = \int_{-\infty}^x N_{m,\sigma}(\gamma) d\gamma$$

C'è una relazione tra le due funzioni di ripartizione:

$$\Phi_{\substack{\text{non} \\ \text{standard}}} \uparrow \Phi_{m,\sigma}(x) = \Phi_{\substack{\text{standard} \\ \text{SF}}} \uparrow \Phi\left(\frac{x-m}{\sigma}\right)$$

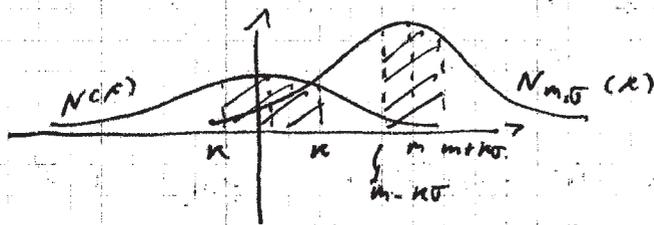
$$\left. \begin{array}{l} \text{vale} \\ \Phi(-x) = \\ 1 - \Phi(x) \end{array} \right\}$$

TEOREMA

$$P(|X_{m,\sigma} - m| \leq k\sigma) = P(|X_{0,1}| \leq k) = 2\phi(k) - 1 \quad \text{ovunque sia } k \geq 0$$

CIOÈ SI PUÒ CALCOLARE LA PROBABILITÀ NON DI UN INTERVALLO A SINISTRA, MA CHE IL NUMERO ALEATORIO CON DISTRIBUZIONE NORMALE m, σ SI SCOSTI DAL SUO VALORE MEDIO DI k VOLTE σ ; SI DIMOSTRA CHE LA PROBABILITÀ DI QUESTO EVENTO È QUELLA CORRISPONDENTE ALLA DISTRIBUZIONE NORMALE STANDARD.

CONSIDERANDO EVENTI SIMMETRICI AL VALORE MEDIO, LA PROBABILITÀ NON DIPENDE DA m O DA σ MA SOLO DA k .



LE DUE AREE SONO UGUALI, L'INTERVALLO È SIMMETRICO RISPETTO DI $\pm k\sigma$

PER σ piccolo si ha una incertezza e si fanno più alti.

GRAZIE A QUESTO TEOREMA TULLO SI RIDUCE ALLA STANDARD.

Dimostrazione: $P(m - k\sigma \leq X \leq m + k\sigma) = \Phi_{m,\sigma}(m + k\sigma) - \Phi_{m,\sigma}(m - k\sigma)$

$$\Phi\left(\frac{m + k\sigma - m}{\sigma}\right) - \Phi\left(\frac{m - k\sigma - m}{\sigma}\right) = \Phi(k) - \Phi(-k)$$

LEB. 16 C.V.S.

TEORIA DELL'AFFIDABILITA'

IL NUMERO ALEATORIO COINVOLTO NEI PROBLEMI DI AFFIDABILITA' E' QUELLO CHE SI CHIAMA IL TEMPO DI ATTESA FINO AL PRIMO QUANTO DI UNA APPARECCHIATURA

X = tempo di attesa fino al primo guasto e un numero aleatorio continuo; il codominio di X ha la potenza del continuo.

In breve, X = "durata"

La funzione di sopravvivenza $S(x)$, che calcola la probabilità dell'evento contrario a quello della funzione di ripartizione di un numero aleatorio X con funzione di ripartizione $F(x)$ e:

$$S(x) = P(X > x) = 1 - F(x)$$

$F(x)$ e' la probabilità che $X \leq x$, $F(x) = P(X \leq x)$

Quindi, se X = "durata" (ad es. in ore) allora

$P(X > t)$ = probabilità che un componente nuovo duri almeno t ore

La probabilità condizionata ha un ruolo fondamentale:

$P(X > s+t | X > s)$ è la probabilità che un componente usato che ha funzionato per s ore duri per almeno altre t ore. Questo ci porta al concetto di fenomeno senza memoria:

$$P(X > s+t | X > s) = P(X > t)$$

FENOMENO SENZA MEMORIA.

Nella teoria dell'affidabilità gioca un ruolo la distribuzione esponenziale:

$$\lambda > 0 \quad f_\lambda(x) = \lambda e^{-\lambda x} \quad \text{per } x > 0$$
$$f_\lambda(x) = 0 \quad \text{per } x \leq 0$$

La funzione di sopravvivenza è così calcolata:

$$F_\lambda(x) = 0 \quad \text{per } x \leq 0$$

p. di ripartizione della
variabile x , $x \leq 0$

$$F_\lambda(x) = \int_0^x \underbrace{f_\lambda(t)}_{\text{densità}} dt = \int_0^x \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x =$$

$P(X \leq x)$

$$= 1 - e^{-\lambda x}$$

Per un assieme $S(x) = e^{-\lambda x}$,

la probabilità che una apparecchiatura la cui durata sia regolata dalla distribuzione esponenziale come densità di probabilità è $e^{-\lambda x}$

La probabilità che un componente usato duri almeno altre t ore è:

$$\begin{aligned} P(X > \lambda + t \mid X > \lambda) &= \frac{P(X > \lambda + t)}{P(X > \lambda)} = \\ &= \frac{S(\lambda + t)}{S(\lambda)} = \frac{e^{-\lambda(\lambda + t)}}{e^{-\lambda\lambda}} = e^{-\lambda t} = S(t) = \\ &= P(X > t) \end{aligned}$$

Il fenomeno è "senza usura", o "senza memoria".

TASSO DI AVARIA

È UN PARAMETRO PER DARE UN SENSO ALL'USURA

$$h(x) = \frac{f(x)}{S(x)} \quad \text{ovvero} \quad \frac{\text{densità}}{f. di sopravvivenza}$$

L'andamento del tasso di avaria determina il tipo di "usura", "usure", "senza memoria":

$h(x)$ crescente \rightarrow usura

$h(x)$ decrescente \rightarrow usure

$h(x)$ costante \rightarrow senza memoria

La distribuzione esponenziale ha tasso di averia costante, $h(x) = \lambda$.

È facile vedere che, una volta assegnata la funzione $h(x)$, è univocamente individuato la densità $f(x)$:

$$(1) f(x) = h(x) e^{-\int_0^x h(t) dt}$$

$$h = \frac{f}{S}, \quad S = 1 - F, \quad F' = f \text{ nel caso continuo}$$

Abbiamo

$$h(x) = \frac{f(x)}{S(x)} = \frac{F'(x)}{1 - F(x)} \text{ che dimostra la (1)}$$

La (1) è importante perché è più facile assegnare il tasso di averia piuttosto che la densità.

Da questo segue una condizione necessaria, assimilabile alla coerenza:

$$S(x) \rightarrow 0 \text{ per } x \rightarrow \infty; \quad F(x) \rightarrow 1 \text{ per } x \rightarrow \infty$$

$$\int_0^{\infty} h(t) dt = +\infty, \text{ allora } h(x) \text{ è un tasso di averia.}$$

Alcuni casi particolari

• $h(x) = \lambda$ costante $\rightarrow f(x) = \lambda e^{-\int_0^x \lambda dt} = \lambda e^{-\lambda x}$

che è la caratterizzazione delle esponenziale.

IL TASSO DI AVARIA È COSTANTE SE E SOLO SE LA DISTRIBUZIONE È ESPONENZIALE.

• $h(x) = \alpha + \beta x$ con α e β costanti distinte, deve valere

$h(x) \geq 0$ per ogni $x \Rightarrow \alpha \geq 0, \beta \geq 0$

($\beta = 0$ si ritrova l'esponenziale e $\alpha > 0$)

$\beta > 0$ usura crescente e la densità è

$f(x) = (\alpha + \beta x) e^{-\alpha x - \frac{1}{2}\beta x^2}$

SE LA FUNZIONE TASSO DI AVARIA È LINEARE ALLORA ESSA È CRESCENTE E RAPPRESENTA UNA USURA

• $h(x) = cx^\beta$ con c, β opposti e $c \neq 0$,

$h(x) \geq 0 \Rightarrow c > 0$

$\int_0^{+\infty} h(t) dt = \int_0^{+\infty} ct^\beta dt = +\infty$ se $\beta > -1$

Espressione della densità: posto $\lambda = \frac{c}{\beta+1}$ si ha, per $x > 0$

$f(x) = \lambda (\beta+1) x^\beta e^{-\lambda x^{\beta+1}}$ che è la densità di Weibull con $\lambda > 0, \beta > -1$

DISTRIBUZIONE DI WEIBULL

È IMPORTANTE NELLA TEORIA DELL'AFFIDABILITÀ.

$$f(x) = \lambda (\beta + 1) x^\beta e^{-\lambda x^{\beta+1}}$$

TASSO DI AVARIA CRESCENTE PER $\beta > 0$ E DECRESCENTE PER $-1 < \beta < 0$; PER $\beta = 0$ ESPONENZIALE

VETTORI ALEATORI

DISTRIBUZIONE DI PROBABILITA' NEL CASO MULTIDIMENSIONALE.

IL VETTORE ALEATORIO m DIMENSIONALE

$$X = (X_1, X_2, \dots, X_m) : \Omega \rightarrow \mathbb{R}^m$$

Funzione da Ω a \mathbb{R}^m ;

Ω famiglia degli eventi; $\mathbb{R}^m \Rightarrow m$ -upla di reali

NEL CASO BIDIMENSIONALE ABBIAMO $(X, Y) : \Omega \rightarrow \mathbb{R}^2$

VETTORE ALEATORIO DISCRETO

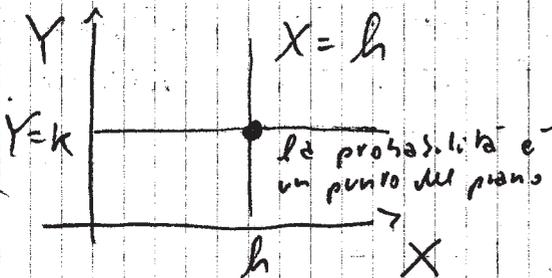
$$P_{hk} = P(X=h, Y=k) > 0$$

IL NUMERO ALEATORIO DISCRETO AVEVA UN NUMERO FINITO

o AL PIU' NUMERABILE DI VALORI POSSIBILI CON

PROBABILITA' POSITIVE E SI ASSEGNANO LE PROBABILITA'

DEI VALORI POSSIBILI $X = x_h$



ESISTONO UN NUMERO FINITO

o AL PIU' UNA INFINITA'

NUMERABILE DI PUNTI

DI COORDINATE

$X = h, Y = k$ SI

AVAN PARE PROBABILITA'

POSITIVA

VETTORE ALEATORIO CONTINUO

Per ogni $(x, y) \in \mathbb{R}^2$,

$$P(X=x, Y=y) = 0$$

$$P(A) = \iint_A f(x, y) dx dy$$

Δ rappresenta
l'insieme
dei punti

$$X=x \text{ e } Y=y$$

Densità di probabilità congiunta,
per fare riferimento alla
distribuzione globale della
coppia X, Y .

DISTRIBUZIONI MARGINALI

NELLA DISTRIBUZIONE BIVARIATA SI TRATTA DI VERDERE
QUAL È IL PROBLEMA DI SOLARE LA DISTRIBUZIONE
AD E). NELLA SOLA X , DATE LE DISTRIBUZIONI
CONGIUNTE, SIANO ESSE CONTINUE O DISCRETE.

Ci si divide della formula di integrazione
sia nel caso discreto, sia nel caso continuo.

• CASO DISCRETO

$$P(X=h) = P(X=h) \cap \Omega =$$

$$= P(X=h) \cap \left[\bigcup_{k \in G} (Y=k) \right] =$$

si presta

$k \in G$

si versa

$$= \sum_k \underbrace{[(X=h) \wedge (Y=k)]}_{\text{equivalente a } (X=h, Y=k)}$$

$$P'_h = P(X=h) = \sum_k P_{hk}$$

LA PROBABILITA' DEL PRIMO COMPONENTE (NON LA DERIVATA) E' LA SOMMA DELLE CONGIUNTE RISPETTO ALL'ALTRA VARIABILE (SI SATURA L'INDICE k SOMMANDO RISPETTO A k)

Quando per trovare la probabilità marginale della sola X si devono sommare le probabilità congiunte rispetto all'altra variabile.

Analogamente, le marginali rispetto a Y è dato da:

$$P''_k = P(Y=k) = \sum_h P_{hk}$$

• CASO CONTINUO

Per fare le marginali nel caso continuo si cambiano le somme in integrali e si rimpiazzano le probabilità concentrate con le densità, più differenziabili.

Si definisce $f_X(x)$ densità marginale di X :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

e, analogamente

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

e, in generale, vale $f(x, y) \neq f_X(x) \cdot f_Y(y)$,
cioè le congiunte e diverse dal prodotto
delle due marginali nel caso continuo.

DISTRIBUZIONI MARGINALI CONDIZIONATE

SI APPLICANO LE LEGGI DELLA PROBABILITÀ CONDIZIONATA

$$P_{h|k}^i = P(X=h | Y=k) = \frac{P_{hk}}{P_k''} \quad \text{per } h=1, 2, \dots$$

con P_{hk} probabilità dell'intersezione congiunta

e P_k'' la marginale rispetto a Y .

$$P_{k|h}'' = P(Y=k | X=h) = \frac{P_{hk}}{P_h^i} \quad \text{per } k=1, 2, \dots$$

Vale la legge del prodotto

$$P_{hk} = P_h^i P_{k|h}'' = P_k'' P_{h|k}^i$$

Per le continue

Se $f_Y(y) > 0$, $f_X(x) > 0$ allora $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$

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

$f(x,y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y)$
per il teorema delle probabilità composte

con $f(x,y)$ densità congiunta

$f_Y(y)$ densità marginale della Y

$f_X(x|y)$ marginale della X condizionata
ad un valore y .

PREVISIONE e VARIANZA

$$U = \varphi(X, Y)$$

numero elettronico

$$IP(U) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x,y) \underbrace{c(x,y)}_{\text{densità congiunta}} dx dy$$

$$\text{var}(U) = IP(U^2) - [IP(U)]^2 \quad \text{densità congiunta}$$

caso particolare

$U = \varphi(X, Y) = X + Y$ vale l'additività della previsione

$$IP(X+Y) = IP(X) + IP(Y)$$

INDIPENDENZA DEI VETTORI ALEATORI

Dati X, Y componenti del vettore aleatorio (X, Y) , si dicono stocasticamente indipendenti se

$$f(x, y) = \alpha(x) \beta(y)$$

per qualunque scelta di x e y .

La previsione del prodotto è

$$IP(XY) = IP(X)IP(Y) \quad \text{vale il prodotto delle previsioni}$$

L'INDIPENDENZA CONSENTE QUINDI DI SCRIVERE LA PREVISIONE DEL PRODOTTO COME IL PRODOTTO DELLE PREVISIONI. QUESTO CI RILIEVA AL CONCETTO DI COVARIANZA:

$$\text{cov}(X, Y) = IP(XY) - IP(X)IP(Y) = 0$$

per due numeri aleatori indipendenti. Non vale il contrario.

Si ricorda che la covarianza può essere scritta così:

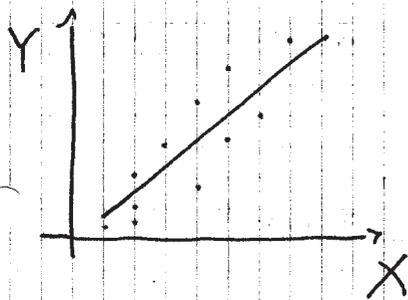
$$\text{cov}(X, Y) = IP(XY) - IP(X)IP(Y) = IP[(X - m_x)(Y - m_y)]$$

prodotto degli scarti dal valore medio

$$m_x = IP(X) \quad m_y = IP(Y)$$

REGRESSIONE

LA RETTA DI REGRESSIONE È LA RETTA CHE PASSA IL PIÙ VICINO POSSIBILE A PUNTI ALEATORI DI \mathbb{R}^2 CON GRANDE PROBABILITÀ.



Dati: (X, Y) vettore aleatorio

$\rho(x, y)$ distribuzione congiunta (densità o probabilità)

valori possibili: punti aleatori di \mathbb{R}^2

La retta di regressione è a coefficienti (costanti) del tipo:

$$y = \alpha + \beta x$$

Per la sua determinazione, prendiamo, in senso verticale, un punto, la cui distanza dalla retta è calcolabile come:

$$\varphi(x, y) = [Y - \alpha - \beta x]^2$$

$\varphi(x, y)$ è funzione di vettore aleatorio

Il quadrato della distanza è per non calcolare con il modulo; la distanza deve essere piccola in valore assoluto, il segno non ha rilevanza.

La distanza stessa è un numero aleatorio che dipende dai parametri α e β .

Calcoliamo la previsione determinando α e β tali che la rendano minima.

$$IP([Y - \alpha - \beta X]^2) =$$

$$= IP([Y - m_2 - \beta(X - m_1) + m_2 - \alpha - \beta m_1]^2) \text{ con } m_1 = IP(X) \text{ e } m_2 = IP(Y), \text{ che sono stati aggiunti e tolti.}$$

Poniamo $\sigma_1^2 = \text{var}(X)$ e $\sigma_2^2 = \text{var}(Y)$:

$$IP([Y - \alpha - \beta X]^2) = \sigma_2^2 + \beta^2 \sigma_1^2 + (m_2 - \alpha - \beta m_1)^2 - 2\beta \text{cov}(X, Y)$$

Per fare il minimo di questa espressione occorre calcolare le derivate parziali rispetto ad α e β , ponendole a zero, e si trova:

$$\alpha = m_2 - \rho \frac{\sigma_2}{\sigma_1} m_1 = \alpha_1 \quad \beta = \rho \frac{\sigma_2}{\sigma_1} = \beta_{21}$$

con ρ che è una covarianza normalizzata, la covarianza diviso il prodotto delle varianze.

Quindi la retta di regressione, di Y in X , è:

$$y = \alpha_1 + \beta_{21} x$$

Se $\rho^2 = 1$ la retta di regressione è il casuale di (Y, X) e tutte le coppie X, Y stanno su una retta, il

codominio del vettore aleatorio (X, Y) è tutto su una retta; la correlazione è lineare.

$$\min IP ([Y - \alpha - \beta X]^2) = \sigma_2^2 (1 - \rho^2) \leq \sigma_2^2$$

Analiticamente, la retta di regressione di X su Y è:

$$X = \alpha_2 + \beta_{12} Y$$

$$\text{con } \beta_{12} = \rho \frac{\sigma_1}{\sigma_2} \quad \text{e} \quad \alpha_2 = m_1 - \rho \frac{\sigma_1}{\sigma_2} m_2$$

Si noti che ρ è simmetrico, scambiando X e Y si scambiano m_1 e m_2 e σ_1 e σ_2 .

Se abbiamo due rette di regressione, la loro eventuale intersezione può rappresentare il baricentro della distribuzione.

Se sono ortogonali ($\rho = 0$) è la situazione peggiore, cioè le componenti del vettore aleatorio sono stocasticamente indipendenti, le due rette sono parallele agli assi coordinati.

Se le due rette coincidono ($\rho^2 = 1$) la situazione è ottimale, con il codominio di (X, Y) .

TEOREMA DI BAYES

per VETTORI ALEATORI

Tale teorema è quello che ricorre $\beta(y|x)$ della n-tupla identifi. che rappresenta due modi per scrivere la distribuzione congiunta:

$$2(x) \beta(y|x) = \beta(y) 2(x|y)$$

(1) $\beta(y|x) = k(x) \beta(y) 2(x|y)$ con $k(x) = \frac{1}{2(x)}$
e, ricordando gli eventi, si ricostruisce il legame fra le componenti del vettore aleatorio.

Il Teorema di Bayes per gli eventi era in fatti un corollario del Teorema delle probabilità composte: dati due eventi E, H ,

$$P(E|H) = \underbrace{P(E)}_{2(x)} \underbrace{P(H|E)}_{\beta(y|x)} = \underbrace{P(H)}_{\beta(y)} \underbrace{P(E|H)}_{2(x|y)}$$

e, per il Teorema di Bayes:

$$P(H|E) = \frac{P(H) P(E|H)}{P(E)} \quad (2)$$

disintegrazione

La (1) e la (2) hanno simmetrie;

$K(x)$, nella (2), si può trovare alle fine perché, nel caso della distribuzione continua, vettori unidimensionali, K è già contenuto nello (1) perché la distribuzione condizionata $\beta(y|x)$, per ogni x fissato è una distribuzione rispetto a y .

Se si tiene x fisso e si fa l'integrale della distribuzione su tutto y , esso deve fare 1 (condizione di normalizzazione di densità).

Faccendo l'integrale anche al secondo membro, K è una costante che viene fuori e questo integrale sotto rispetto a y di questa quantità è proprio $Q(x)$, e $Q(x)$ di integrato.

IL CAMPIONAMENTO STATISTICO

Esso sta alla base dell'inferenza statistica che si divide in teoria della stima e verifica dell'ipotesi.

Lo strumento utilizzato sarà il Teorema di Bayes per vettori aleatori.

$$\beta(\theta | x) = K(x) \beta(\theta) \theta^{E_{xi}} (1-\theta)^{n-E_{xi}}$$

con

$$K(x) = \frac{1}{\int_0^1 \beta(\theta) \alpha(x|\theta) d\theta}$$

} Formula di interpretazione del continuo.

Una rilevazione su due campionario ottenendo i risultati per fare inferenza ossia per dedurre dal campione stesso un'informazione relativa all'intera popolazione.

IL CAMPIONAMENTO STATISTICO CASUALE

L'URNA RAPPRESENTA UNA POPOLAZIONE STATISTICA, COSTITUITA DA INDIVIDUI CHE POSSONO AVERE O NO UN DATO ATTRIBUTO O CARATTERISTICA. E SI CERCA, INFORMARSI

L'ESPERIMENTO COMPATENTE IN SUCCESSIVE ESTRAZIONI CASUALI, DI ACQUISIRE INFORMAZIONI SULLA FRAZIONE ϑ DI INDIVIDUI CON DIFESA CARATTERISTICA

Ad ogni estrazione di un individuo della popolazione si può associare il numero aleatorio X_i , indicatore dell'evento E_i : ma la distribuzione di probabilità di X_i non è determinata, perché dipende dal parametro ϑ .

In effetti, dal numero aleatorio X_i (che può assumere solo valori 0 o 1) si conosce la distribuzione condizionata ad ogni valore ϑ di ϑ , che si può scrivere

$$f(x_i | \vartheta) = \begin{cases} \vartheta & \text{se } x_i = 1 \\ 1 - \vartheta & \text{se } x_i = 0 \end{cases}$$

che è detta distribuzione di Bernoulli.

Ad un n ϑ è frazione di polline bianca in unurna di composizioni ignota, all'incertezza del risultato dell'estrazione (E_i = bianca alla i -esima estrazione) si aggiunge quella su ϑ , che può essere il valore di un numero aleatorio ϑ con codominio $[0, 1]$ e conviene supporre ϑ continuo, con densità opportunamente assegnata. La probabilità (condizionata a ϑ) di successo (estrazione di polline bianca) è $P(E_i | \vartheta)$.

Considerando n estrazioni, n numeri
 elettori X_1, X_2, \dots, X_n individuano il
 vettore elettorale $X = (X_1, X_2, \dots, X_n)$,
 componente del vettore (X, Θ) ; se si indica
 con α $(x_1, x_2, \dots, x_n | \Theta)$ o, in breve, con $\alpha(x | \Theta)$,
 la distribuzione marginale di X condizionale
 a Θ , si ha (μ "molto minore" del numero
 totale di individui nella popolazione):

$$\alpha(x | \Theta) = f(x_1 | \Theta) f(x_2 | \Theta) \dots f(x_n | \Theta) =$$

$$= \Theta^{\sum x_i} (1 - \Theta)^{n - \sum x_i}$$

μ $n = 1 \dots n$ che
 rappresenta il numero
 dei successi.

Se si designa anche la
 distribuzione marginale $\beta(\Theta)$ di Θ , μ la
 distribuzione congiunta di (X, Θ) si ha:

$$c(x, \Theta) = \beta(\Theta) \alpha(x | \Theta)$$

e quindi la distribuzione marginale di X è:

$$\alpha(x) = \int_0^1 c(x, \Theta) d\Theta = \int_0^1 \beta(\Theta) f(x_1 | \Theta) f(x_2 | \Theta) \dots f(x_n | \Theta) d\Theta$$

che è una versione "continua" di
 la successi in n prove con restituzioni da una di
 composizione α nota.

Applicando (con ϑ al posto di γ) il teorema di Bayes si può determinare la distribuzione puntuale (casi condizionate e x) di ϑ , che è, usando la precedente espressione di $z(x|\vartheta)$:

$$\beta(\vartheta|x) = k(x) \beta(\vartheta) \vartheta^{\sum x_i} (1-\vartheta)^{n-\sum x_i}$$

con $k(x) = 1 / \int_0^1 \beta(\vartheta) z(x|\vartheta) d\vartheta$

CAMPIONAMENTO DA DISTRIBUZIONE NORMALE

INTERFERA UNA CARATTERISTICA RAPPRESENTATA DA UN NUMERO ALEATORIO X CON DISTRIBUZIONE NORMALE.

Distribuzione normale, densità $N_{\mu, \sigma^2}(x)$, nota la varianza σ^2 (precisione $\frac{1}{\sigma^2} = h$) ma non il valore medio μ .

Si considera la distribuzione iniziale:

$$\beta(\vartheta) = N_{\mu_0, \sigma_0}(\vartheta)$$

Questa scelta rappresenta una sintesi dello stato di informazione (precedente alla osservazione del campione), del tipo:

(ii) è molto probabile che il valore di θ sia compreso nell'intervallo $(m_0 - 3\sigma_0, m_0 + 3\sigma_0)$;

(iii) la stessa probabilità che θ appartenga a uno qualunque dei due intervalli di uguale ampiezza e simmetrica rispetto a m_0 ;

(iv) in ogni intervallo del tipo $(m_0 - k\sigma_0, m_0 + k\sigma_0)$ con $k > 0$ arbitrario, il valore centrale sono il "più probabile" e quello a "meno"

Se si considera una sola osservazione allora la funzione di verosimiglianza è:

$$l(x|\theta) = N_{\theta, \sigma_0}(x)$$

e quindi μ il teorema di Bayes

$$\beta(\theta|x) = K(x) N_{m_0, \sigma_0}(\theta) N_{\theta, \sigma_0}(x)$$

e a calcoli fatti, otteniamo una distribuzione finale, quella condizionata a x .

$$\beta(\theta|x) = N_{m_n, \sigma_n}$$

$$\text{con } m_n = \frac{\frac{n}{\sigma^2} \bar{x} + \frac{1}{\sigma_0^2} m_0}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

m_n è una media ponderata delle m_0 iniziali e delle medie aritmetiche, con pesi uguali alle precisioni.

\bar{x} è la media aritmetica

m_0 è l'opinione iniziale

el denominatore la somma dei pesi

$$\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$$

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

INTERVALLI

DI CONFIDENZA

Dato α (ad es. $\alpha = 0.95$) e introducendo un sottoinsieme A del codominio di \mathcal{X} tale che

$P(\mathcal{U} \in A | x) = \alpha$, così che la probabilità che \mathcal{U} appartenga ad A condizionata ad x per chi questa cosa si fa dopo l'osservazione; quindi tutti le probabilità in gioco sono sempre condizionate.

Vogliamo che la probabilità sia molto grande:

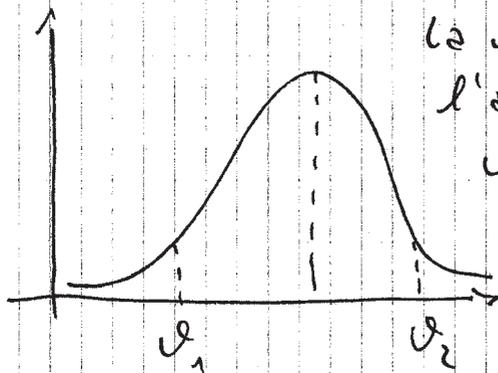
$$P(\vartheta \in A | X) = \int_A \beta(\vartheta | X) d\vartheta = 2$$

\uparrow
incognita

LA PROBABILITÀ DI APPARTENERE AD UN INTERVALLO È L'INTERVALLO ESTREMO ALL'INTERVALLO DELLA DISTRIBUZIONE, DELLA $\beta(\vartheta | X)$.

$A = [\vartheta_1, \vartheta_2]$ intervallo di confidenza al 100 α %.

Inoltre, una volta fatto il campionamento siamo arrivati ad una distribuzione finale, migliore di quella iniziale.



La distribuzione finale;
l'area è la probabilità che ϑ sia compreso tra ϑ_1 e ϑ_2 , ad es. 0.95.

Formulario

Coefficiente di correlazione	$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$
Covarianza	$\text{Cov}(X, Y) = P(XY) - P(X)P(Y)$
Densità di probabilità	$f_X(x)$
— condizionata	$f_X(x y) = \frac{f(x, y)}{f_Y(y)}$
— congiunta	$f(x, y)$
— marginale	$f_X(x) = \int f(x, y) dy$
Distribuzione	
— binomiale	$P(X = x) = \binom{n}{x} p^x q^{n-x}; x = 0, 1, 2, \dots, n$
— discreta	$p_X(x)$
— — condizionata	$p_X(x y) = \frac{p(x, y)}{p_Y(y)}$
— — congiunta	$p(x, y)$
— — marginale	$p_X(x) = \sum_{y \in C_Y} p(x, y)$
— esponenziale	$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{per } x \geq 0 \\ 0 & \text{per } x < 0. \end{cases}$

— finale

$$\beta(\theta | x) = \frac{a(x | \theta)\beta(\theta)}{\int a(x | \theta)\beta(\theta) d\theta};$$

$\beta(\theta) =$ distribuzione iniziale

— gamma

$$G_{c,\lambda}(x) = \begin{cases} \frac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

— geometrica

$$P(X = x) = pq^{x-1}; x = 1, 2, 3, \dots$$

— ipergeometrica

$$P(X = x) = \frac{\binom{b}{x} \binom{N-b}{n-x}}{\binom{N}{n}};$$

$$\max(0, b + n - N) \leq x \leq \min(b, n)$$

— normale

$$N_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right];$$

$-\infty < x < +\infty$

— Pascal

$$P(X = x) = \binom{x-1}{k-1} p^k q^{x-k};$$

$x = k, k+1, k+2, \dots$

— Poisson

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}; x = 0, 1, 2, \dots$$

— Rayleigh

$$f(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right] & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

— uniforme

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{per } x \in [a, b] \\ 0 & \text{altrove} \end{cases}$$

Evento certo

Ω

Evento impossibile

\emptyset

Funzione

— gamma

$$\Gamma(c) = \int_0^{+\infty} x^{c-1} e^{-x} dx; c > 0$$

— generatrice delle probabilità

$$\mathbb{P}(t^X) = \sum p_n t^n; n \geq 0$$

— generatrice dei momenti

$$\mathbb{P}(e^{tX})$$

— di ripartizione	$F(x) = \int_{-\infty}^x f(t) dt$
— di ripartizione normale standard	$\Phi(x) = \int_{-\infty}^x N_{0,1}(t) dt$
— di sopravvivenza	$S(x) = 1 - F(x)$
— di verosimiglianza	$\begin{aligned} a(x \theta) &= a(x_1, x_2, \dots, x_n \theta) \\ &= f(x_1 \theta)f(x_2 \theta)\cdots f(x_n \theta) \end{aligned}$
Indicatore di evento	$ E = \begin{cases} 1 & \text{se } E \text{ è vero} \\ 0 & \text{se } E \text{ è falso} \end{cases}$
Indipendenza stocastica di A da B	$P(A B) = P(A)$
Indipendenza stocastica di due numeri aleatori X e Y	$c(x, y) = a(x)\beta(y) \quad \forall (x, y) \in \mathbb{R}^2$
Intensità	(vedi tasso di avaria)
Lacune di un processo uniforme: densità	$l(x) = \begin{cases} \frac{n}{c} \left(1 - \frac{x}{c}\right)^{n-1} & \text{per } x \in [0, c] \\ 0 & \text{altrove} \end{cases}$
Media aritmetica	$m = \frac{x_1 + x_2 + \dots + x_N}{N}$
Media armonica	$\frac{1}{\mu} = \frac{1}{N} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_N} \right)$
Previsione	$\begin{aligned} \mathbb{P}(X) &= \sum p_i x_i \quad (X \text{ discreto}); \\ \mathbb{P}(X) &= \int x f(x) dx \quad (X \text{ continuo}) \end{aligned}$
Processo di Poisson	$P[N(t) = k] = \frac{(at)^k}{k!} e^{-at}$
Retta di regressione di Y su X	$\begin{aligned} y &= a_{YX} + \beta_{YX} x; \\ \beta_{YX} &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, a_{YX} = \mathbb{P}(Y) - \beta_{YX} \mathbb{P}(X) \end{aligned}$
— di X su Y	$\begin{aligned} x &= a_{XY} + \beta_{XY} y; \\ \beta_{XY} &= \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}, a_{XY} = \mathbb{P}(X) - \beta_{XY} \mathbb{P}(Y) \end{aligned}$

Scambiabilità di eventi

$$P(E_{i_1} \cap \dots \cap E_{i_k} \cap E_{i_{k+1}}^c \cap \dots \cap E_{i_n}^c) = \\ P(E_{j_1} \cap \dots \cap E_{j_k} \cap E_{j_{k+1}}^c \cap \dots \cap E_{j_n}^c) \\ \text{essendo } (i_1, i_2, \dots, i_n) \text{ e } (j_1, j_2, \dots, j_n) \text{ due qualsiasi} \\ \text{permutazioni degli indici } (1, 2, \dots, n) \text{ (} 0 \leq k \leq n \text{)}$$

— di numeri aleatori

$$c(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = c(x_{j_1}, x_{j_2}, \dots, x_{j_n}), \\ \text{essendo } (i_1, i_2, \dots, i_n) \text{ e } (j_1, j_2, \dots, j_n) \text{ due qualsiasi} \\ \text{permutazioni degli indici } (1, 2, \dots, n)$$

Sigma (σ) additività

$$P(\Omega) = P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) = 1; \\ \{E_i\} = \text{partizione numerabile di } \Omega$$

Tasso di avaria

$$h(x) = \frac{f(x)}{S(x)} \quad (X > 0)$$

Teorema delle probabilità
composte

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

Teorema delle probabilità
totali

$$P\left(\bigcup_{i=1}^k E_i\right) = \sum_{i=1}^k P(E_i) \quad (E_1, E_2, \dots, E_k \text{ eventi incom-} \\ \text{patibili)}$$

Teorema di Bayes

$$P(H_i | E) = \frac{P(E | H_i)P(H_i)}{\sum_{j=1}^N P(E | H_j)P(H_j)} \\ (i = 1, 2, \dots, N)$$

Varianza

$$\text{Var}(X) = \mathbb{P}[(X - M)^2]; \quad M = \mathbb{P}(X)$$

Ottobre - Novembre 2014