STATISTICA

By Edoardo

PRINCIPIO FONDAMENTALE CALCOLO COMBINATORIO

Se una scelta può essere fatta in r modi diversi, per ciascuno dei quali una seconda scelta può essere effettuata in s modi diversi ecc. allora la successione di tutte le scelte può essere compiuta in $r \cdot s \cdot \ldots$ modi diversi.

PERMUTAZIONI SEMPLICI

Dati n oggetti le permutazioni sono il # di combinazioni dell'ordine di tali oggetti
$\Rightarrow n!$

- CON RIPETIZIONI

le combinazioni dell'ordine di n oggetti con h oggetti uguali, k oggetti uguali .... con $h+k+\ldots = n$

$$P^r_{n,h,k\ldots} = \frac{n!}{h!\cdot k!\cdot\ldots}$$

DISPOSIZIONI SEMPLICI DI CLASSE K

Dati n oggetti, le disposizioni sono i raggruppamenti di $k<n$ elementi tutti distinti e ogni raggruppamento differisce da un altro per almeno un elemento o l'ordine.

Il numero di tali disposizioni è dato da:

$$D_{n,k} = \frac{n!}{(n-k)!}$$

- CON RIPETIZIONE

la ripetizione sta nel fatto che in un gruppo di k elementi posso avere fino a k uguali $D^r_{n,k} = n^k$

NB. dato che sono ammesse ripetizioni k può essere anche maggiore di n.

COMBINAZIONI SEMPLICI

Esattamente come le disposizioni con l'unica differenza che i gruppi differiscono solo per almeno un elemento e non per l'ordine

ES. $(A,B) = (B,A)$

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

- CON RIPETIZIONE

Esattamente come con le disposizioni, la ripetizione sta nella possibilità di avere fino a k elementi uguali e k può essere maggiore di n

$$C^r_{n,k} = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$$

BINOMIO DI NEWTON

$$(a+b)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a^{n-k} b^k$$

CAP 3

- Approccio Soggettivo: la probabilità di un evento non è oggettiva, ma la fiducia che lo studioso ripone nel suo verificarsi.
- Approccio Frequentista: la probabilità è una proprietà dell'evento stesso e si pensa possa essere ricavata ripetendo più volte l'esperimento.

SPAZIO DEGLI ESITI E EVENTI

Lo spazio degli esiti (spazio campionario) è l'insieme di tutti gli esiti possibili e si denota con S o Ω
Gli eventi (sottoinsiemi dello spazio campionario) sono insiemi i cui elementi sono esiti possibili.

- $E \cup F$ l'unione di due eventi è un evento che si verifica se si verifica o l'evento E o l'evento F.

TEOREMA PROBABILITÀ TOTALI

Dato un insieme di eventi trovare la probabilità che se ne verifica almeno uno (unione)

- INCOMPATIBILI $\Rightarrow P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
- QUALSIASI: USARE QUESTO

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

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

- $E \cap F$ l'intersezione di due eventi è un evento che si verifica se si verificano sia E che F

TEOREMA PROBABILITÀ COMPOSTA

: La probabilità di un evento composto $(A \cap B)$ è uguale al prodotto della probabilità di un evento per la probabilità dell'altro condizionata al primo.

$P(E \cap F) = P(E) \cdot P(F)$ SE INDIPENDENTI

$P(E \cap F) = P(E) \cdot P(F|E)$ SE DIPENDENTI

NB. Un evento E si verifica se almeno uno dei suoi elementi si verifica.
Se $E \cap F = \emptyset$ ovvero non possono verificarsi entrambi si dicono EVENTI DISGIUNTI
$E^c \in S$ si verifica solo se non si verifica $E \in S \Rightarrow$ EVENTO COMPLEMENTARE

TEOREMA PROBABILITÀ ASSOLUTA

Dato l'insieme di eventi U lo dividiamo in sottoinsiemi $H_1, H_2 \dots$
Ogni evento sarà anch'esso partizionato $A = (A \cap H_1) + (A \cap H_2) \dots$
e così la sua probabilità $P(A) = P(A \cap H_1) + P(A \cap H_2) \dots$
Applicando il teorema della probabilità composta si ha: $P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_1) \dots$
$\Rightarrow P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i)$ ⊢ FORMULA DI DISINTEGRAZIONE

DISUGUAGLIANZA DI BOOLE

Per ogni insieme finito o numerabile di eventi la probabilità che accada almeno un evento è minore o uguale alla somma delle probabilità dei singoli eventi.

$$P(\bigcup_i A_i) \leq \sum_i P(A_i)$$

PROBABILITÀ CONDIZIONATA

Probabilità di un evento dato un evento precedente

$P(E|F)$ - probabilità di E dato F.

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

TEOREMA DI BAYES

Data una quantità finita di eventi $F_1, F_2 \dots F_n$ mutualmente esclusivi e un evento E verificato:

$$P(F_J|E) = \frac{P(F_J \cap E)}{P(E)} = \frac{P(E|F_J)P(F_J)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

$$E = \bigcup_{i=1}^{n} (E \cap F_i)$$

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

INDIPENDENZA EVENTI

Due eventi sono indipendenti se $P(A \cap B) = P(A) \cdot P(B)$

Dato l'evento C con $P(C) \neq 0$ due eventi si dicono condizionatamente indipendenti da C se $P(A \cap B|C) = P(A|C) \cdot P(B|C)$

ASSIOMI STATISTICA

1) La probabilità di un evento è compresa tra 0 e 1
2) La probabilità dell'evento certo è $1 \rightarrow P(\Omega) = 1$
3) $P(A \cup B) - P(A) + P(B)$ se indipendenti

CAP 4

VARIABILI ALEATORIE

È una variabile che può racchiudere tutti i possibili esiti di un esperimento.
Es. lancio di due dadi, E sarà la variabile aleatoria della somma.
$\Rightarrow P(E=3)$ = probabilità che la somma sia $3 = \{(1,2),(2,1)\} = 2/36$

FUNZIONE DI RIPARTIZIONE

La funzione di ripartizione della variabile aleatoria X è definita per ogni numero reale x tramite:

$F(x) = P(X \leq x)$ la funzione $F(x)$ esprime la probabilità che X sia $\leq x$
$(X \sim F)$

Tutte le questioni di probabilità che ammettono variabili aleatorie ammettono anche una risposta in termini di della funzione di ripartizione.

– V.A. DISCRETE

Una v.a. si dice discreta se può assumere una quantità finita o numerabile di valori.
Data una variabile aleatoria discreta si definisce FUNZIONE DI MASSA $P(a) = P(X=a)$
$P(a) > 0$ per tutti gli esiti possibili di X $\;\longrightarrow\; \sum_{i=1}^{\infty} P(x_i) = 1$
$P(a) = 0$ per tutti gli altri valori

La funzione di ripartizione può essere espressa in funzione della funzione di massa.

$$F(a) = \sum_{x \leq a} P(x)$$

– V.A. CONTINUE

Una variabile aleatoria che possa assumere un'infinità non numerabile di valori si dice continua

$P(X \in B) = \int_B \boxed{f(x)} dx$ FUNZIONE DI DENSITÀ

Tale equazione dice che la probabilità che X appartenga a un insieme B si può trovare integrando la sua densità su tale insieme.

$$1 = P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x)\,dx$$

NB. Tutte le probabilità che riguardano una variabile continua possono essere espresse in termini di integrali dell'integrale della sua densità

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

$P(X = a) = \int_a^a f(x)\,dx = 0$

La densità è la derivata della funzione di ripartizione

FUNZIONE DI RIPARTIZIONE CONGIUNTA

Necessaria per conoscere il rapporto tra due variabili aleatorie dello stesso esperimento.

$$F(x,y) = P(X \leq x \cap Y \leq y)$$

$$\Rightarrow F_X(x) = P(X \leq x) = P(X \leq x \cap Y < \infty) = F(x \cap \infty)$$

$$F_Y(y) = F(\infty \cap y)$$

FUNZIONE DI MASSA CONGIUNTA

$$P(x_i, y_j) = P(X = x_i \cap Y = y_j) \qquad i = 1, 2 \dots \quad j = 1, 2 \dots$$

Le funzioni di massa individuali si possono ricavare:

$$P_X(x_i) = P\left[\bigcup_j (X = x_i \cap Y = y_j)\right] = \sum_j P(x_i, y_j)$$

$$\hookrightarrow P_Y(y_j) = \sum_i P(x_i, y_j)$$

FUNZIONE DI DENSITÀ CONGIUNTA

Due variabili aleatorie X e Y sono congiuntamente continue se esiste una funzione non negativa $f(x,y)$ avente la proprietà che per ogni sottoinsieme C del piano cartesiano

$$P((X,Y) \in C) = \iint_{(x,y) \in C} f(x,y)\,dx\,dy$$

$$P(X \in A \cap Y \in B) = \int_B \int_A f(x,y)\,dx\,dy$$

Ponendo $A = (-\infty, a]$ e $B = (-\infty, b]$ la formula di ripartizione diventa:

$$F(a,b) = P(X \leq a \cap Y \leq b) = P(X \in A \cap Y \in B) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\,dx\,dy$$

$$\Rightarrow f(a,b) = \frac{d^2 F(a,b)}{da\,db}$$

Le densità marginali si ricavano:

$$\cancel{\int_A} f_X(x)\,d\cancel{x} = P(X \in A) = P(X \in A \cap Y \in \mathbb{R}) = \cancel{\int_A} \int_{-\infty}^{\infty} f(x,y)\,dy\,d\cancel{x}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \vee \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx \qquad \left[\frac{1}{\text{area}}, \ \underset{\text{altrove}}{0}\right]$$

Due variabili aleatorie dello stesso esperimento si dicono indipendenti se tutti gli eventi relativi ad una sono indipendenti da tutti quelli relativi all'altra.

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

In caso contrario si dicono dipendenti.
Dall'indipendenza deduciamo anche:
$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b) \Rightarrow F(a,b) = F_X(a)F_Y(b)$ (ripartizione $\leq x$)
Se in più sono anche discrete:
$P(x,y) = P_X(x)P_Y(y)$ la funzione di massa congiunta è uguale al prodotto delle marginali. (massa $= x$)
Nel caso invece che X e Y siano congiuntamente continue esse sono indipendenti solo se la densità congiunta è il prodotto delle marginali.

$$f(x,y) = f_X(x) f_Y(y)$$

Questi concetti si estendono anche in caso avessimo n variabili:
la funzione di ripartizione congiunta sarà $F(a_1, a_2, \dots a_n) = P(X \leq a_1, X \leq a_2 \dots, X \leq a_n)$
se sono discrete la funzione di massa congiunta sarà
$P(x_1, x_2, \dots x_n) = P(X = x_1, X = x_2 \dots, X = x_n)$
mentre se sono congiuntamente continue la densità congiunta sarà
$P(X_1 \in A_1, X_2 \in A_2, \dots X_n \in A_n) = \int_{A_1} \int_{A_2} \dots \int_{A_n} f(x_1, x_2 \dots x_n) dx_1 dx_2 \dots dx_n$

$$E[X] = \int_a^b f_X(x)dx \wedge E[Y] = \int_a^b f_Y(y)dy$$

VALORE ATTESO

Il valore atteso di una variabile aleatoria è il numero $E[X] = \sum_i x_i p(X = x_i)$ e corrisponde alla media pesata dei valori possibili di X usando come pesi le probabilità che X assuma tali valori. (media o aspettazione)

— VALORE ATTESO DI UNA FUNZIONE DI VARIABILE ALEATORIA

$$E[g(X)] = \sum_x g(x)p(x) \quad \text{DISCRETA}$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx \quad \text{CONTINUA}$$

— $E[aX + b] = aE[X] + b$

Nel caso bidimensionale:

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y) & \text{DISCRETE} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy & \text{CONTINUE} \end{cases}$$

— $E[X+Y] = E[X] + E[Y]$ (Discrete e Continue)

$\Rightarrow E[X_1 + X_2 \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$

— $E[XY] = E[X]E[Y]$ se indipendenti

VARIANZA

$Var(X) = E[X^2] - E[X]^2$ esprime la misura della variabilità dei valori assunti da x, nello specifico di quanto si discostino quadraticamente dal valore atteso.

La radice quadrata della varianza $\sqrt{Var(X)}$ è detta deviazione standard della variabile X

- $Var(aX + b) = a^2 Var(X)$
- $Var(X + b) = Var(X)$ $\Rightarrow Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$ Se indipendenti
- $Var(aX) = a^2 Var(X)$

NB. $Var(X) = E[(X - E[X])^2]$

$Cov(X,Y) = E[XY] - E[X]E[Y]$ è detta COVARIANZA

Essa esprime quanto X e Y variano insieme cioè quanto sono dipendenti.

- $Cov(X,Y) = Cov(Y,X) \Rightarrow Cov(X,Y) = 0$ se indipendenti
- $Cov(X,X) = Var(X)$
- $Cov(aX,Y) = a\,Cov(X,Y) = Cov(X,aY)$
- $Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$ proprietà additiva

$$\hookrightarrow cov\left(\sum_{i=1}^{n} X_i, Y\right) = \sum_{i=1}^{n} cov(X_i, Y)$$

$$\hookrightarrow cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} cov(X_i, Y_j)$$

Tenendo a mente le prime due proprietà:

$$Var\left(\sum_{i=1}^{n} X_i\right) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) =$$

$$= \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} Cov(X_i, X_j)$$ Varianza di somma di v.a.

Nel caso di $n=2$

$$Var(X+Y) = Var(X) + Var(Y) + 2\,cov(X,Y)$$

LEGGE DEBOLE DEI GRANDI NUMERI

- DISUGUAGLIANZA DI MARKOV

$$P(X \geq a) \leq \frac{E[X]}{a}$$ con X v.a. mai negativa e $a > 0$

DIMOSTRAZIONE:

$$E[X] = \int_0^{\infty} x f(x)\,dx = \underbrace{\int_0^{a} x f(x)\,dx}_{A} + \int_a^{\infty} x f(x)\,dx$$

$$\geq \int_a^{\infty} x f(x)\,dx \geq \int_a^{\infty} a f(x)\,dx = a \int_a^{\infty} f(x)\,dx = a\,P(X \geq a)$$

Poichè A positivo — poichè $x \geq a$

$\Rightarrow E[X] \geq a\,P(X \geq a)$ divido entrambi per a e ottengo la disuguaglianza.

- Disuguaglianza di Chebyshev

$$P(|x-\mu| \geq r) \leq \frac{\sigma^2}{r^2}$$ con $\mu = E[x]$ e σ = variazione standard = $\sqrt{Var(x)}$

Dimostrazione:

$|X-\mu| \geq r$ è equiprobabile a $(x-\mu)^2 \geq r^2$ ed avendo sicuramente nel secondo caso una v.a. positiva possiamo applicare la disuguaglianza di Markov con $a = r^2$

$$P(|X-\mu| \geq r) = P((X-\mu)^2 \geq r^2) \leq \frac{\overbrace{E[(X-\mu)^2]}^{Var(X)}}{r^2} = \frac{\sigma^2}{r^2}$$

Ponendo $r = k\sigma$ si ha $P(|X-\mu| > k\sigma) \leq \frac{1}{k^2}$ ovvero, la probabilità che una variabile aleatoria differisca dalla sua media per più di $k \cdot \sqrt{Var(x)}$ volte non può superare $1/k^2$

Dalla definizione di queste due disuguaglianze arriviamo alla legge debole dei grandi numeri:

Data $X_1, X_2 \dots$ una successione di variabili aleatorie i.i.d. (indipendenti e identicamente distribuite) tutte con $E[X_i] = \mu$. Allora per ogni $\varepsilon > 0$

$$P\left(\left|\frac{X_1+X_2 \dots + X_n}{n} - \mu\right| > \varepsilon\right) \longrightarrow 0 \qquad n \to \infty$$

Dimostrazione:

- $Var(X_i) = \sigma^2$ • $E\left[\frac{X_1+X_2 \dots + X_n}{n}\right] = \mu$ • $Var\left(\frac{X_1 + \dots X_n}{n}\right) = \frac{\sigma^2}{n}$

Dalla disuguaglianza di Chebyshev applicata alla v.a. $(X_1 \dots + X_n)/n$ avremo:

$$P\left(\left|\frac{X_1 + \dots X_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$ dato che il secondo membro tende a zero per n che tende a infinito, l'enunciato è provato.

CAP. 5

VARIABILI ALEATORIE BERNOULLI E BINOMIALI

• Una v.a. si dice bernoulliana se la sua funzione di massa è:

$P(X=0) = 1-p$ $\qquad E[X] = p$
$P(X=1) = p$

• Una variabile si dice binomiale di parametri (n,p) se denota il numero totale di successi con probabilità p in n esperimenti distinti; la sua funzione di massa sarà:

$$P(X=i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad Var(X) = n \cdot p(1-p)$$

Se X_1 e X_2 sono binomiali di parametri (n_1, p) e (n_2, p) indipendenti allora $X_1 + X_2$ è una binomiale di parametri $(n_1 + n_2, 1$

FUNZIONE DI RIPARTIZIONE: $P(X \leq i) = \sum_{k=0}^{i} \binom{n}{p} p^k (1-p)^{n-k}$

• $P(X=k+1) = P(X=k) \frac{p}{1-p} \cdot \frac{n-k}{k+1}$

VARIABILI ALEATORIE DI POISSON

Una variabile aleatoria si dice di Poisson se assume solo valori non negativi ed esprime la probabilità che si verifichino n eventi in un dato tempo sapendo che in media se ne verificano λ. La sua distribuzione è:

$$P(X=n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

$$E[X] = Var(X) = \lambda$$

Se una binomiale ha n molto grande e p molto piccolo può essere considerata di Poisson e avrà la sua stessa funzione di massa e $E[X] = \lambda = n \cdot p$

$$P(X=0) = \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda}$$

$$\Rightarrow P(X=i+1) = \frac{\lambda}{i+1} P(X=i)$$

VARIABILI ALEATORIE IPERGEOMETRICHE

$$P(X=i)=\frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}$$

$i = \#$ successi

$N =$ casi favorevoli $\quad n -$ ripetizioni

$M =$ casi sfavorevoli

Una variabile aleatoria che ha questa funzione di massa è detta ipergeometrica con parametri N, M e n

Data $X_i = \begin{cases} 1 & \text{successo} \\ 0 & \text{altrove} \end{cases} \qquad n=1$

$$P(X_i=1)=\frac{N}{N+M}$$

$$E[X_i]=p=\frac{N}{N+M} \quad \Rightarrow X=\sum X_i \Rightarrow E[X]=n\frac{N}{N+M}=n \cdot p$$

$$Var(X_i)=\frac{NM}{(N+M)^2} \quad \Rightarrow \quad // \quad \Rightarrow np(1-p)\left[1-\frac{n-1}{N+M-1}\right]$$

VARIABILI UNIFORMI

Una v.a. continua si dice uniforme nell'intervallo $[\alpha,\beta]$ se ha funzione di densità:

$$f(x)=\begin{cases} \frac{1}{\beta-\alpha} & \alpha \leq x \leq \beta \\ 0 & \text{altrimenti} \end{cases}$$

$$P(X\in C) = \frac{1}{\beta-\alpha}\int_a^b dx = \frac{b-a}{\beta-\alpha} \qquad (C=[a,b] \in [\alpha,\beta])$$

$$E[X]=\int_\alpha^\beta \frac{x\,dx}{\beta-\alpha}=\frac{\alpha+\beta}{2}$$

$( )^2$

$$Var(X)=E[X^2]-E[X]^2=\frac{(\beta-\alpha)^2}{12}$$

VARIABILI NORMALI O GAUSSIANA

Una v.a. X si dice normale o gaussiana di parametri (μ,σ^2) e si scrive $X\sim N(\mu,\sigma^2)$ se X ha funzione di densità:

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\cdot e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

$$E[X]=\mu \qquad Var(X)=\sigma^2$$

N.B. Una trasformazione lineare di una gaussiana è anch'essa una gaussiana.

Funzioni Generatrici dei Momenti

- $g(t) = E[e^{tX}] = \sum_{i=1}^{n} p_i e^{tx_i}$ $\quad p =$ funzione di massa

- $g(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$ $\quad f(x) =$ funzione di densità

- $g'(0) = E[X]$
- $g''(0) = E[X^2]$

Data $X \sim N(\mu, \sigma^2)$ e $Y = aX + b$ con a, b reali costanti e $a \neq 0$:
Y è una gaussiana con $E[Y] = a\mu + b$ e $Var(Y) = a^2\sigma^2$

Se $Z = \frac{X-\mu}{\sigma} \Rightarrow E[Z] = 0$ e $Var(Z) = 1$ $\quad Z$ è gaussiana standard

e la sua funzione di ripartizione $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{-y^2/2}\,dy$

$$P(Z < a) = \phi \qquad P(Z > a) = 1 - \phi$$

Variabili Aleatorie Esponenziali

Una v.a. continua si dice esponenziale con parametro $\lambda > 0$ se la sua funzione di densità è:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad F(x) = \int_0^x f(x)\,dy = 1 - e^{-\lambda x}$$

$E[X] = 1/\lambda \quad Var(X) = 1/\lambda^2$

La distribuzione esponenziale rappresenta il tempo che passa prima di un certo evento dall'istante t.
La proprietà centrale della distribuzione esponenziale è l'assenza di memoria:

$$P(X > s+t \mid X > t) = P(X > s)$$

$$= \frac{P(X > s+t,\ X > t)}{P(X > t)}$$

$\Rightarrow P(X > s) \cdot P(X > t) = P(X > s+t)$ perché se $X > s+t$ allora $X > t$

DISTRIBUZIONI CHE DERIVANO DA QUELLA NORMALE

• CHI-QUADRO

Se $Z_1, Z_2 \dots Z_n$ sono v.a. normali standard indipendenti allora la somma dei loro quadrati:
$X = Z_1^2 + Z_2^2 \dots + Z_n^2$ è una v.a. che prende il nome di chi-quadro a n gradi di libertà $\Rightarrow X \sim X_n^2$

$P(X \geq X_{\alpha,n}^2) = \alpha \qquad \alpha \in (0,1) \qquad X_{\alpha,n}^2 = \dfrac{1}{\Phi(1-\alpha)}$

• DISTRIBUZIONI t

Date Z normale standard e C_n chi-quadro con n gradi di libertà

$\Rightarrow T_n = \dfrac{Z}{\sqrt{C_n/n}}$ = distribuzione t $\Rightarrow T_n \sim t_n$

$E[T_n] = 0 \quad n \geq 2 \qquad P(T_n \geq t_{\alpha,n}) = \alpha \qquad \alpha \in (0,1)$

$Var[T_n] = \dfrac{n}{n-2} \quad n \geq 3$

• DISTRIBUZIONI F

Dati C_n e C_m indipendenti allora la v.a. $F_{n,m}$ definita da:

$F = \dfrac{C_n/n}{C_m/m}$ si dice avere distribuzione F con n e m gradi di libertà.

$P(F_{n,m} > F_{\alpha,n,m}) = \alpha \qquad \alpha \in (0,1)$

CAP. 6

STATISTICA CAMPIONARIA

Un insieme $X_1, X_2 \dots X_n$ di v.a. indipendenti tutte con la stessa distribuzione F si dice campione o campione aleatorio della distribuzione F

MEDIA CAMPIONARIA

$$\overline{X} = \frac{X_1 + X_2 + \dots X_n}{n}$$

anch'essa è una variabile aleatoria e di conseguenza ha media e varianza

$E[\overline{X}] = E[X_i]$

$Var(\overline{X}) = \frac{Var(X_i)}{n}$

TEOREMA LIMITE CENTRALE

Siano $X_1, X_2 \dots X_n$ delle v.a.i.i.d. con media μ e varianza σ^2. Se n è molto grande la somma $Y = X_1 + X_2 + \dots + X_n$ è approssimativamente normale con media $n\mu$ e varianza $n\sigma^2$

Y può essere normalizzata $\Rightarrow Z = \frac{Y - n\mu}{\sigma\sqrt{n}}$

$\Rightarrow P(Z < x) = \Phi(x)$

NB. $\Phi(x)$ necessita delle tavole

APPROSSIMAZIONE MEDIA CAMPIONARIA

$\overline{X} = 1/n \sum_{i=1}^{n} x_i$ $\Rightarrow$ se n è grande $\overline{X}$ è approssimativamente normale per il teorema sopra con media $\mu = E[X_i]$ e deviazione standard $\sigma/\sqrt{n} = \sqrt{Var(\overline{X})}$

$\Rightarrow$ Normalizziamo: $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

$$P(Z < x \wedge Z > -x) = P(Z < x)$$

VARIANZA CAMPIONARIA

Dati $X_1, X_2 \dots X_n$ con media μ e varianza σ^2, $\overline{X}$ la media campionaria:

S^2 = varianza campionaria $= \underbrace{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}$ e S = variazione campionaria standard

$E[S^2] = \sigma^2$

$$\rightarrow \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2\right) \frac{n}{n-1} = \frac{\sum x^2 - n\overline{x}^2}{n-1}$$

Se $X_1, X_2 \dots X_n$ è un campione proveniente da una distribuzione normale di media μ e varianza σ^2 allora $\overline{X}$ e S^2 sono indipendenti. Inoltre $\overline{X}$ è normale con media μ e varianza σ^2/n, mentre $(n-1)S^2/\sigma^2$ è una chi-quadro con n-1 gradi di libertà.

- $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ = distribuzione t con n-1 gradi di libertà.

CAP. 7

In statistica a differenza di quanto accade in probabilità le distribuzioni in gioco non sono note completamente. pertanto facciamo dell'inferenza sui parametri sconosciuti usando i dati osservati.

STIMATORI DI MASSIMA VEROSIMIGLIANZA

Uno stimatore di parametro θ è una variabile aleatoria che ha lo scopo di dare una stima di tale parametro basandosi su altre informazioni.

Dati $X_1 \dots X_n$ denotiamo con $f(x_1, \dots x_n \mid \theta)$ la loro funzione di densità/massa condizionata al valore θ.
Tale funzione esprime la "verosimiglianza" che si realizzi la n-pla quando θ è proprio il valore dell'incognita: quello che cerchiamo di dare è la stima di θ che rende massima la verosimiglianza ovvero $f(x_1 \dots x_n \mid \theta)$ anche detta funzione di likelihood.
Gli stimatori che hanno il compito di dare tale stima sono chiamati stimatori di verosimiglianza.

$f(x_1 \dots x_n \mid \theta)$ e $\log[f(x_1 \dots x_n \mid \theta)]$ vengono massimizzate per lo stesso valore di θ. La seconda è detta funzione di log-likelihood.

- Stimatore di massima verosimiglianza della probabilità di successo di una distribuzione Bernoulliana

$$d(X_1 \dots X_n) = 1/n \sum_{i=1}^{n} X_i$$

- Stimatore di massima verosimiglianza di λ di una distribuzione Poissoniana

$$d(X_1 \dots X_n) = 1/n \sum_{i=1}^{n} X_i$$

- Stimatori di massima verosimiglianza della media μ e variazione standard σ di una distribuzione normale

1) $\overline{X}$ 2) $\sqrt{\frac{1}{n} \sum_i (X_i - \overline{X})^2}$

- Stimatore di massima verosimiglianza per la media di una distribuzione uniforme.

$$\frac{\max(X_1 \dots X_n)}{2} \qquad (0, \theta)$$

INTERVALLI DI CONFIDENZA DISTRIBUZIONE NORMALE

— Rispetto la media

• Varianza nota

Dato un campione di una popolazione normale con media μ incognita e varianza σ^2 nota sappiamo che lo stimatore di massima verosimiglianza della media è $\bar{X} = \Sigma X_i/n$, ma questo non vuol dire che la media abbia proprio quel valore, ma che sia vicina ad esso. Stabiliamo quindi un intervallo di fiducia o confidenza dato da:

$$\left(\bar{X} - z_{\frac{1-\%}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{1-\%}{2}} \frac{\sigma}{\sqrt{n}}\right) \in \mu$$

Precisione Stima: ampiezza intervallo $\Rightarrow$ differenza estremi.

Affermiamo così che con il tot% di confidenza la media vera appartiene a quell'intervallo.

• Entrambi incogniti

L'unica differenza è che necessitiamo di calcolare la varianza campionaria

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

Come prima $\mu \in \left(\bar{X} - t_{\frac{1-\%}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{1-\%}{2}} \frac{S}{\sqrt{n}}\right)$

NB. L'unica differenza è che invece di z vi è t poiché:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

da trovare sulla tabella di Student

— Rispetto la varianza

• Entrambi incogniti

$$\sigma^2 \in \left(\frac{(n-1)S^2}{X^2_{\frac{1-\%}{2}, n-1}}, \frac{(n-1)S^2}{X^2_{\frac{1+\%}{2}, n-1}}\right)$$

NB. Il valore delle chi-quadro $X^2_{a,n}$ è da trovare sulle tabelle.

Livello di confidenza	90%	95%	98%	99%
α	0.10	0.05	0.02	0.01
$Z_{\alpha/2}$	1.645	1.96	2.33	2.58

CAP. 8

VERIFICA DELLE IPOTESI

Data una popolazione con distribuzione F_θ che dipende da θ incognito vogliamo verificare l'ipotesi su θ detta ipotesi nulla H_0. Se F_θ ha varianza 1 e media θ per esempio possiamo avere:

1) $H_0: \theta = 1$
2) $H_0: \theta \leq 1$

Nel primo caso abbiamo $N(1,1)$ che se vera caratterizza l'intera distribuzione e quindi è detta ipotesi semplice
Nel secondo caso invece no e quindi è detta ipotesi composta

L'ipotesi H_0 su un campione $(X_1 \dots X_n$ è accettata se tale campione $\notin C$ dove C è la regione di spazio di n dimensioni che corrisponde al test (regione critica del test). Rifiutata altrimenti.

Nel caso in cui l'ipotesi H_0 viene rifiutata perché incompatibile, ma in realtà risulta vera si parla di errore di prima specie. Al contrario se ne accettiamo una che in realtà è falsa si parla di errore di seconda specie.

La probabilità di errore di un test deve essere $\leq \alpha$ dove α corrisponde al livello di significatività.

– Varianza nota, media incognita

$H_0: \mu = \mu_0$ (valore richiesto) $\qquad H_1: \mu \neq \mu_0$

H_1 è accettata se $\left|\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}\right| > z_{\frac{\alpha}{2}}$, rifiutata altrimenti

La probabilità p-dei-dati fornisce il livello di significatività critico di H_0:

$$P\left(|Z| > \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}\right) = \text{p-dei-dati} = 2P\left(Z > \left|\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}\right|\right)$$

L'ipotesi viene accettata per ogni $\alpha <$ p-dei-dati

$\beta(\mu)$ è la probabilità degli errori di seconda specie (detta curva OC) quando μ è la media reale

$$\beta(\mu) = \phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\frac{\alpha}{2}}\right) - \phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\frac{\alpha}{2}}\right)$$

dove μ_0 = media ipotetica e μ = media reale.

$1-\beta(\mu)$ è detta funzione di potenza del test che corrisponde alla probabilità di rifiutare correttamente H_0 con μ media reale.

La curva OC permette di dimensionare il campione affinché l'errore di seconda specie soddisfi condizioni specifiche

ES. $H_0: \mu = \mu_0$ con valore reale $\mu = \mu_1$

$\beta(\mu_1) = c$

$$\Rightarrow n = \left[\frac{(z_{\frac{a}{2}} + z_{\beta})\sigma}{\mu_1 - \mu_0}\right]^2$$

• Test bilaterali

– Destro:

$\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ p-dei-dati $= P\left(z > \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}\right)$

H_1 viene accettata se $\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > z_a$, rifiutata altrimenti

– Sinistro

$\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_1 \end{cases}$ p-dei-dati $= P\left(z < \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}\right)$

H_1 si accetta se $\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < -z_a$

NB. a = livello di significatività

– Varianza e media incognite

$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ p-dei-dati $= 2(1 - \phi(|t|))$

H_1 si accetta se $\left|\frac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| > t_{\frac{a}{2}, n,1}$

$\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$ p-dei-dati $= \phi(t)$

H_1 si accetta se $\frac{\overline{X} - \mu_0}{\sqrt{n}} < -t_{a, n-1}$

$\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ p .-dati $= 1 - \phi(t)$

H_1 si accetta se $\frac{\overline{X} - \mu_0}{S/\sqrt{n}} > t_{a, n,1}$